
MOLECULAR CHARACTERIZATION OF AROMATIC COMPOUND AND HEAVY METAL DETOXIFICATION SYSTEMS IN THERMOPHILIC MICROORGANISMS: POTENTIAL IMPACT ON BIOMONITORING AND BIOREMEDIATION

Immacolata Del Giudice

Dottorato in Scienze Biotecnologiche – XXV ciclo
Indirizzo Biotecnologie Industriali e Molecolari
Università di Napoli Federico II



Dottorato in Scienze Biotecnologiche – XXV ciclo
Indirizzo Biotecnologie Industriali e Molecolari
Università di Napoli Federico II



**MOLECULAR CHARACTERIZATION
OF AROMATIC COMPOUND AND
HEAVY METAL DETOXIFICATION
SYSTEMS IN THERMOPHILIC
MICROORGANISMS: POTENTIAL
IMPACT ON BIOMONITORING AND
BIOREMEDIATION**

Immacolata Del Giudice

Dottoranda: Immacolata Del Giudice
Relatore: Prof.ssa Simonetta Bartolucci
Coordinatore: Prof. Giovanni Sannia

'Le nostre valigie logore stavano di
nuovo ammucchiate sul marciapiede;
avevano altro e più lungo cammino da
percorrere. Ma non importa, la strada è vita.'

Jack Kerouac

Contents

Riassunto	1
Summary	7
Chapter 1	General introduction 8
Chapter 2	Identification and physicochemical, structural and functional characterization of BldR2 from <i>Sulfolobus solfataricus</i> P2 37
Chapter 3	Structural and functional characterization of TtArsC, a novel arsenate reductase involved in arsenic detoxification in the bacterium <i>Thermus thermophilus</i> HB27 59
Chapter 4	Characterization of a putative transcriptional regulator, TtSmtB, involved in arsenic detoxification in the bacterium <i>Thermus thermophilus</i> HB27 79
Chapter 5	General conclusion 94
List of talks and publications	99

RIASSUNTO

La diffusione nell'ambiente di sostanze tossiche, tra le quali arsenico e composti aromatici, ed i loro effetti dannosi sulla salute umana rappresentano un serio problema in molte aree del mondo. In Bangladesh, India, Vietnam e Nord America, per esempio, milioni di persone sono esposte ogni giorno a livelli di arsenico nelle acque potabili che oltrepassano i limiti imposti dalla 'WHO' ('World Health Organization'). Siti contaminati da composti aromatici sono diffusi in tutto il mondo, per esempio, in Germania sono state individuate più di 1400 aree, deputate all'estrazione carbonifera, contaminate da idrocarburi monoaromatici, policiclici aromatici e composti fenolici; solo una piccola percentuale (20%) di queste aree è stata presa in considerazione per opere di bonifica.

L'arsenico è un metalloide normalmente presente nell'ambiente, ma le attività umane hanno contribuito al suo accumulo anomalo nella biosfera e, di conseguenza, al suo inserimento nella catena alimentare dove ha provocato gravi danni a tutti gli organismi viventi. Anche i composti aromatici hanno un'origine naturale, in quanto possono derivare dalla degradazione di polimeri come la lignina, tuttavia, dall'avvento dell'industrializzazione ad oggi, la loro concentrazione ambientale è andata sempre più aumentando a causa del loro impiego nell'industria farmaceutica e chimica, nei combustibili fossili e nei pesticidi.

Dato che alcuni di questi composti presentano una tossicità molto elevata, appare evidente che è richiesto un attento monitoraggio dell'inquinamento dei suoli e delle acque, sia per adottare politiche di prevenzione che azioni di 'bioremediation', soprattutto in quelle zone altamente contaminate localizzate in prossimità di aree industriali o in cui tali aree sono state dismesse. Negli ultimi anni, al posto delle tradizionali metodiche chimiche per la determinazione della concentrazione ambientale di inquinanti, ha trovato largo impiego l'utilizzo dei biosensori cellulari come sistema di monitoraggio ambientale. Rispetto alle tecniche convenzionali, questi ultimi offrono diversi vantaggi, quali la semplicità d'uso, la specificità, la velocità, l'economicità e la possibilità di effettuare misure tossicologiche. I biosensori cellulari sono microrganismi, geneticamente ingegnerizzati, dotati, solitamente, di un plasmide contenente una fusione genica tra un promotore, sensibile alla sostanza da monitorare, ed un gene *reporter*, codificante una proteina la cui espressione può essere facilmente determinata. I geni *reporter* più frequentemente utilizzati sono quelli codificanti la 'green fluorescent protein' (GFP), la β -galattosidasi o la luciferasi, che possono essere rivelate mediante fluorescenza, saggi colorimetrici e bioluminescenza, rispettivamente. Di recente è stato riportato l'utilizzo di un nuovo gene *reporter* (*crtA*), codificante una proteina coinvolta nella biosintesi dei carotenoidi, che permetterebbe di costruire biosensori cellulari basati su metodi di rivelazione ad occhio nudo.

Ad oggi, il microrganismo più utilizzato come biosensore cellulare è stato *Escherichia coli* e pochi tentativi sono stati fatti per usare microrganismi diversi ed, in ogni caso, hanno sempre previsto l'utilizzo di mesofili. Data la loro resistenza a condizioni ambientali estremamente sfavorevoli, un passo avanti nel *design* di sensori batterici più stabili e più efficienti potrebbe venire dall'impiego di microrganismi estremofili.

A causa della resistenza delle loro molecole alle alte temperature, unita a quella ai solventi organici, detergenti, estremi di pH ed agenti caotropici, tra i microrganismi estremofili, quelli che hanno suscitato il maggiore interesse sono stati i termofili e gli ipertermofili.

Al fine di caratterizzare i sistemi di detossificazione di composti aromatici e arsenico ed individuare molecole termostabili e termoresistenti, da poter utilizzare nel biomonitoraggio e nella *bioremediation* di tali composti, in questo lavoro di ricerca sono stati utilizzati come organismi modello due microrganismi termofili: *Sulfolobus solfataricus* P2 e *Thermus thermophilus* HB27.

Nella prima parte di questo lavoro è stata posta particolare attenzione alla caratterizzazione del *pathway* di detossificazione dei composti aromatici (in particolare benzaldeide e salicilato) e della sua regolazione in *S. solfataricus*. In uno studio condotto precedentemente era già stato caratterizzato un meccanismo, basato su un sistema trasportatore/regolatore, implicato nella detossificazione della benzaldeide in questo organismo; in particolare, il regolatore BldR, appartenente alla famiglia di regolatori trascrizionali MarR, è stato individuato come un attivatore trascrizionale che, in presenza di benzaldeide, è in grado di attivare l'espressione di una permeasi, capace di estrarre la benzaldeide all'esterno della cellula, e di una alcool deidrogenasi, capace di mitigarne la tossicità andando a convertire la benzaldeide nel corrispondente alcool.

In questo lavoro di tesi, è stata caratterizzata una nuova proteina, denominata BldR2, anche esso appartenente alla famiglia MarR. Tali regolatori sono stati ampiamente studiati al fine di comprendere le basi molecolari dei meccanismi di resistenza a droghe e della risposta allo stress indotto da composti aromatici. Il regolatore identificato, BldR2, è stato espresso in forma ricombinante in *E. coli*, purificato all'omogeneità e caratterizzato strutturalmente e funzionalmente. La proteina si è dimostrata capace di legarsi specificamente, come omodimero, alla sequenza promotrice a monte del suo stesso gene; mediante un'analisi della sequenza nucleotidica, unita ad esperimenti di DNAsi-I *footprinting*, è stato individuato, nella regione di legame, un sito costituito da una sequenza pseudo palindromica di 8 coppie di basi separata da 3 nucleotidi da una palindrome perfetta di 8 coppie di basi. Dato che questo sito si sovrappone agli elementi basali identificati del promotore e che, da dati riportati in letteratura, i siti localizzati in questa regione sono solitamente le sequenze riconosciute dai repressori trascrizionali, è stato ipotizzato che BldR2 possa regolare la sua stessa espressione andando ad interferire con il legame dei fattori basali (TBP and TFB). Inoltre, è stato dimostrato che la proteina è capace di legare alcuni composti aromatici, come benzaldeide e salicilato, con affinità micromolari e che, in seguito al legame con essi, perde l'affinità verso la sua sequenza promotrice *target* in una maniera dipendente dalla concentrazione. Salicilato e benzaldeide agiscono come effettori di diversi membri MarR, anche se non è escluso che ligandi naturali a più alta affinità possano non essere stati ancora individuati.

I risultati *in vitro* ottenuti per BldR2 sono stati in accordo con l'induzione *in vivo* dell'espressione del gene *bldR2* in seguito all'aggiunta di composti aromatici.

Dato che i geni codificanti molti membri MarR sono localizzati a monte o a valle dei geni regolati, è stato studiato l'ambiente genomico di *bldR2* e sono stati individuati due geni, localizzati a monte di *bldR2*, codificanti due componenti di un sistema di trasporto di peptidi antimicrobici. Mediante saggi EMSA è stata dimostrata la capacità di BldR2 di legare il promotore a monte di questi geni, così come quello a monte del gene *adh*, codificante l'alcool deidrogenasi, e dell'operone MarR-like, codificante la proteina BldR e la permeasi, precedentemente citate. Il legame a questi promotori avviene con affinità confrontabili a quella verso il promotore del suo stesso gene. Dai dati emersi è stato possibile ipotizzare un ruolo di BldR2 nella regolazione della risposta allo stress indotto da composti aromatici, anche se ulteriori

esperimenti sono richiesti per chiarire il ruolo dei due regolatori MarR, BldR e BldR2, che condividono alcune sequenze di DNA *target*. Dato che è stato dimostrato che l'espressione del gene *bldR2* avviene in presenza dei composti aromatici in una fase tardiva della crescita di *S. solfataricus* rispetto a quella di *bldR*, che è principalmente espresso nella fase esponenziale iniziale, è stato ipotizzato che BldR2 nella cellula potrebbe coadiuvare o complementare la funzione di BldR nella regolazione dell'espressione dei geni in seguito, per esempio, ad un accumulo dei composti induttori oltre un valore soglia. La ridondanza funzionale è, del resto, molto riportata in letteratura. Un'altra ipotesi potrebbe riguardare, invece, una modulazione dell'attività di BldR2 da parte di effettori endogeni derivanti da *pathway* catabolici aromatici.

Questa ricerca potrebbe avere differenti applicazioni biotecnologiche, soprattutto nel campo della costruzione di biosensori cellulari, che possono essere disegnati usando i promotori sensibili ai composti aromatici identificati in questo studio. Ulteriori analisi sono richieste per ottenere una migliore conoscenza del meccanismo globale di regolazione di BldR2, in particolare per quel che riguarda l'identificazione dei geni *target* e dei ligandi naturali, in modo da poter costruire sistemi sensori dotati di sequenze "più sensibili" agli inquinanti in esame rispetto a quelle attualmente in commercio.

La seconda parte di questo lavoro di ricerca è consistita nella caratterizzazione del meccanismo di resistenza all'arsenico e della sua regolazione in un altro microrganismo termofilo: *T. thermophilus* HB27. Questo batterio si è dimostrato capace di sopravvivere in presenza di elevate concentrazioni di arsenito ed arseniato, le due forme di arsenico più diffuse nell'ambiente. L'elevata resistenza all'arsenico ha fatto ipotizzare la presenza in questo microrganismo di geni coinvolti nella detossificazione di tale metalloide.

In diversi microrganismi, i geni per la resistenza all'arsenico sono stati individuati sia sui plasmidi che sui cromosomi di molti batteri e sono stati esaustivamente caratterizzati. Tali geni sono solitamente organizzati in operoni di tre, *arsRBC*, o cinque, *arsRDABC*, geni codificanti: un regolatore trans-agente ArsR, il quale funziona come un repressore che controlla i livelli di espressione dei geni *target* in risposta alla presenza di arsenito, una arseniato reduttasi citoplasmatica, ArsC, in grado di convertire l'arseniato nel più tossico arsenito, una proteina con attività ATPasica, ArsA, che, insieme alla pompa di efflusso ArsB, forma un complesso in grado di estrarre l'arsenito attraverso la membrana citoplasmatica, e la proteina ArsD, per cui non è stata ancora attentamente caratterizzata la funzione, che sembrerebbe facilitare il legame dell'arsenito al complesso ArsAB.

In questo studio sono stati identificati e caratterizzati due componenti del sistema di detossificazione *ars* di *T. thermophilus*: una arseniato reduttasi, TtArsC, e un regolatore trascrizionale appartenente alla famiglia di regolatori ArsR/SmtB, TtSmtB. Sono stati, inoltre, individuati due geni codificanti putativi trasportatori ArsB, un gene codificante un trasportatore con un motivo di legame ai metalli pesanti (HMA) ed altri tre geni codificanti putativi regolatori trascrizionali della famiglia ArsR/SmtB. Dato che per alcuni organismi appartenenti al genere *Thermus* è stata riportata la capacità di ossidare l'arsenito, grazie alla presenza di un arsenito ossidasi, e di utilizzare l'arseniato come accettore finale di elettroni nella respirazione anaerobica (sistema dei geni *arr*), si è voluta investigare la presenza di tali abilità anche in *T. thermophilus* HB27. Le analisi genomiche non hanno evidenziato la presenza di alcuna putativa arsenito ossidasi o di alcun putativo gene del sistema *arr* responsabile del metabolismo dell'arseniato. Gli unici individuati sono stati i geni *ars*, in un

arrangiamento insolito rispetto a quello degli altri organismi in cui il sistema è stato caratterizzato; essi, infatti, non sono organizzati in un singolo operone *ars*. Il primo determinante della resistenza all'arsenico ad essere stato caratterizzato in questo lavoro è *TtArsC*. L'analisi trascrizionale di questo gene ha rivelato la presenza del trascritto in ogni condizione testata (presenza o meno di arsenito e arseniato) e la sua trascrizione in un operone con i due geni immediatamente a monte codificanti un'ipotetica proteina ed un'oligoendopeptidasi. Inoltre, esperimenti di qRT-PCR hanno dimostrato che i livelli di espressione del gene sono molto bassi ma incrementano fino a quattro volte in seguito ad esposizione ad arseniato. L'arsenito sembra non sortire alcun effetto sull'espressione del gene *TtarsC*. La proteina *TtArsC* è stata espressa in forma ricombinante in *E. coli*, purificata all'omogeneità e caratterizzata strutturalmente e funzionalmente. Mediante esperimenti di gel filtrazione e *light scattering*, è stato dimostrato che la proteina in condizioni native è presente in forma monomerica. Essa è, inoltre, in grado di ridurre l'arseniato ad arsenito in saggi enzimatici allestiti *in vitro*. Dato che le arseniato reduttasi della famiglia a cui appartiene *TtArsC* richiedono per il loro funzionamento un sistema *redox* in grado di rigenerare il potere riducente dell'enzima, è stato messo a punto un saggio in cui è stato utilizzato un sistema eterologo e ricombinante costituito dalla tioredossina di *E. coli* (Trx) e dalla tioredossina reduttasi di *S. solfataricus* (SsTr). Il saggio è stato seguito monitorando il decremento di assorbanza a 340 nm del NADPH, consumato nella reazione dalla tioredossina reduttasi. L'enzima *TtArsC* si è mostrato effettivamente in grado di ridurre l'arseniato ad arsenito e di usare gli elettroni provenienti dal sistema *redox* Trx/SsTr.

Il meccanismo catalitico della classe di arseniato reduttasi di cui potenzialmente fa parte *TtArsC*, è stato ben caratterizzato: tre cisteine formano il sito catalitico responsabile dell'attività reduttasica, la prima cisteina è quella che effettua l'attacco nucleofilo sull'arseniato, formando un intermedio covalente con esso, successivamente, la seconda cisteina forma un ponte disolfurico con la prima cisteina, determinando il rilascio dell'arsenito. Per rigenerare il primo tiolo, la terza cisteina forma un ponte disolfurico con la seconda rendendo la prima cisteina libera di effettuare un nuovo attacco nucleofilo. I gruppi tiolici delle due cisteine impegnate nel ponte disolfurico sono, poi, rigenerati da una cascata *redox* che coinvolge una tioredossina, una tioredossina reduttasi e, infine, il NADPH.

Allo scopo di caratterizzare tale meccanismo in *TtArsC*, è stato costruito un mutante in cui il primo residuo di cisteina (Cys7) è stato sostituito con un residuo di serina. I saggi enzimatici effettuati con tale mutante hanno mostrato la sua incapacità nel ridurre l'arseniato ad arsenito, come osservato dal mancato consumo nel tempo del NADPH. Si è potuto, quindi, concludere che la cisteina in posizione 7 è un residuo fondamentale per l'attività reduttasica di *TtArsC* e, sicuramente, è il residuo che effettua l'attacco nucleofilo sull'arseniato.

Dato che *TtArsC* presenta omologia con la arseniato reduttasi caratterizzata in *Staphylococcus aureus* e che quest'ultima, oltre all'attività di riduzione dell'arseniato, presenta anche attività fosfatasica, si è voluto indagare se anche *TtArsC* possedesse tale attività. I saggi effettuati col substrato sintetico p-nitrofenil fosfato hanno mostrato che *TtArsC* possiede una debole attività fosfatasica, confermando la sua evoluzione dalle 'low molecular weight protein phosphatases', già riportata per altre arseniato reduttasi caratterizzate. È molto interessante notare che *TtArsC* è la prima arseniato reduttasi proveniente da un batterio Gram⁻ a presentare questa duplice attività (reduttasica/fosfatasica), che è, invece, tipica delle arseniato reduttasi dei batteri Gram⁺; così come è una delle poche arseniato reduttasi da Gram⁻ a presentare un

meccanismo catalitico simile a quello delle arseniato reduttasi da Gram⁺. L'arseniato reduttasi caratterizzata dal batterio Gram⁻ *E. coli*, infatti, presenta un meccanismo catalitico differente da quello precedentemente descritto; quest'enzima non presenta le tre cisteine catalitiche conservate ma ne possiede una sola, usa come potenziale riducente glutatione e glutaredossina e non presenta attività fosfataseica.

Un'altra caratteristica unica di TtArsC è la sua termostabilità, la sua temperatura di *melting*, infatti, è risultata essere di 91°C ed è stato anche dimostrato che la proteina conserva la sua attività catalitica anche dopo essere stata incubata per 90 minuti a 80°C. Queste caratteristiche rendono TtArsC l'arseniato reduttasi più termostabile descritta fino ad ora.

Al fine di valutare la regolazione dei geni per la resistenza all'arsenico in *T. thermophilus*, si è scelto di caratterizzare il putativo regolatore trascrizionale, TtSmtB, il quale, tra i putativi regolatori ArsR/SmtB identificati dalle analisi bioinformatiche, mostrava la più alta identità di sequenza con i membri caratterizzati della famiglia di regolatori sensibili ai metalli ArsR/SmtB. Preliminari analisi della sequenza proteica di TtSmtB hanno mostrato la presenza di un sito di legame all'arsenito, ELCV(C/G)D, altamente conservato. Anche in questo caso, le analisi trascrizionali hanno mostrato che il gene *TtsmtB* è espresso sia nelle cellule sottoposte a stress da arsenico che in quelle non trattate, e che esso è trascritto in un singolo operone con i due geni immediatamente a monte e i due immediatamente a valle, i quali codificano proteine che, apparentemente, non sembrano essere funzionalmente correlate. L'unico tra questi geni che costituiscono l'operone a poter avere un ruolo nella resistenza all'arsenico è il gene *TTC0354*, codificante un putativo trasportatore di cationi che, dalla presenza di un motivo HMA (*heavy metal associated*), lascerebbe supporre un suo possibile ruolo nel trasporto di metalli. L'analisi quantitativa dell'espressione genica di *TtsmtB* ha mostrato che essa viene incrementata di due volte sia in presenza di arseniato che di arsenito. La risposta ad entrambe le forme di arsenico è ampiamente riportata per i membri della famiglia ArsR/SmtB. Sorprendentemente, gli stessi esperimenti effettuati sul gene *TTC0354* hanno mostrato un *pattern* di induzione dell'espressione genica differente; infatti, essa è stata indotta da arsenito (fino a quattro volte) ma non dall'arseniato. Una possibile spiegazione dei diversi livelli di espressione di *TtsmtB* e *TTC0354* potrebbe risiedere nella presenza di un promotore interno all'operone localizzato a monte del gene *TTC0354*, che, in particolari condizioni (per esempio in presenza di elevate concentrazioni intracellulari di arsenito), potrebbe incrementare l'espressione del trasportatore che, a sua volta, potrebbe coadiuvare i trasportatori ArsB nell'estruzione dell'arsenito. La presenza di promotori interni in operoni 'sensibili' ai metalli è stata recentemente riportata da Napolitano *et al.* che hanno dimostrato la presenza di quattro promotori interni nell'operone 'sensibile' allo zinco del cianobatterio *Anaebena* PCC7120, riconosciuti con diversa affinità e da differenti regolatori trascrizionali.

La proteina TtSmtB è stata espressa in forma ricombinante e purificata all'omogeneità. Non è stato possibile effettuare una completa caratterizzazione strutturale e funzionale di TtSmtB a causa della sua instabilità e della sua tendenza all'aggregazione e alla precipitazione. Molti tentativi di stabilizzare la proteina (aggiunta di DTT per mantenere le cisteine nello stato ridotto, aggiunta di stabilizzanti come il TMAO) sono stati effettuati. Tale instabilità non è nuova per i membri della famiglia ArsR/SmtB ed è stata riportata per molti di questi regolatori come, ad esempio, QacR di *S. aureus*. Preliminari saggi EMSA hanno mostrato la capacità della proteina di legare sequenze di DNA, contenenti siti palindromici, localizzate nelle regioni a monte dell'operone che contiene il gene *TtsmtB*, e a monte del gene

TtsmtB stesso. È stata dimostrata, inoltre, la capacità della proteina di legare una regione di DNA localizzata a monte dell'operone *TTC1500/TTC1501/TTC1502*, contenente il gene *TtarsC* precedentemente descritto, contenente il putativo promotore. Per valutare il ruolo dell'arsenito nella modulazione dell'attività di TtSmtB, sono stati effettuati dei saggi EMSA in cui la proteina è stata incubata col promotore *target* in presenza di concentrazioni crescenti di arsenito. Tale metalloide si è dimostrato capace di diminuire l'affinità della proteina verso il suo putativo *target*. La parziale caratterizzazione del ruolo funzionale della proteina TtSmtB ha reso difficile ipotizzare un modello per spiegare il meccanismo molecolare di resistenza all'arsenico in *T. thermophilus*.

In conclusione, i dati raccolti rivelano un significativo contributo delle proteine TtArsC e TtSmtB nella risposta all'arsenico. L'arseniato, una volta entrato nella cellula mediante i trasportatori del fosfato, potrebbe incrementare il livello di espressione di *TtarsC*, probabilmente attraverso l'azione coordinata di TtSmtB e/o di altri regolatori trascrizionali non ancora identificati. Una volta che l'arseniato è stato ridotto ad arsenito, questo potrebbe modulare l'azione del regolatore TtSmtB in modo da garantire una maggiore espressione delle proteine coinvolte nell'estruzione dell'arsenito tossico, sia quello proveniente dalla riduzione dell'arseniato che quello che entra nella cellula dall'esterno mediante le acquagliceroporine.

La conoscenza globale delle basi molecolari e genetiche del meccanismo di detossificazione dell'arsenico, oltre che essere interessante da un punto di vista evolutivo, rappresenta un buon punto di partenza per lo sviluppo di strategie di 'bioremediation' più efficienti e selettive. Infatti, una arseniato reduttasi termofila, o una sua variante con termostabilità o proprietà catalitiche migliorate mediante tecniche di *directed evolution*, potrebbe essere impiegata in combinazione con materiali capaci di intrappolare l'arsenito, per eliminare tale metalloide dagli ambienti contaminati.

SUMMARY

Both arsenic and aromatic compounds are naturally present in the environment but human activities, such as the chemical and pharmaceutical industries, use of fossil fuels and pesticides, have contributed to their anomalous accumulation in the biosphere, determining severe damages to all living organisms. Many microorganisms possess tuned mechanisms for sensing the level of pollutants in their growth environment and controlling intracellular concentrations according to their biochemical needs. In this PhD thesis a mechanism for aromatic compound detoxification and a strategy for arsenic resistance have been identified and characterized in *Sulfolobus solfataricus* P2 and in *Thermus thermophilus* HB27, respectively. In *S. solfataricus* we characterized BldR2, as a new member of the MarR transcriptional factor family, and reported the physiological, biochemical, and biophysical investigation of its stability and DNA binding ability. Transcriptional analysis revealed the upregulation of *bldR2* expression by aromatic compounds and allowed the identification of *cis*-acting sequences. BldR2 is a dimer in solution and possesses a high stability against temperature and chemical denaturing agents; the protein binds site-specifically to its own promoter and the alcohol dehydrogenase gene and the MarR-like operon promoters, as well as to the putative promoter of the operon encoding components of antimicrobial peptide transport system, located immediately upstream of its gene. Benzaldehyde and salicylate, the ligands of BldR2, are antagonists of DNA binding. Two single-point mutants of BldR2 have been produced and characterized; the results point to arginine 19 as a key amino acid involved in protein dimerization, while the introduction of a serine in position 65 increases the DNA affinity of the protein, making it comparable with those of other members of the MarR family. Regarding to the arsenic resistance mechanism, *T. thermophilus* exhibited a good tolerance to high concentrations of arsenate and arsenite; it owns in its genome a putative chromosomal arsenate reductase (*TtarsC*) gene, encoding a protein homologous to the one well characterised from the plasmid pI258 of the Gram⁺ *Staphylococcus aureus* bacterium, and a putative chromosomal transcriptional regulator (*TtsmtB*) gene, encoding a protein homologous to the members of ArsR/SmtB family. Differently from the characterized arsenic resistance genes of many microorganisms, *TtarsC* and *TtsmtB* are part of two operons including genes not apparently related to arsenic resistance; qRT-PCR showed that *TtarsC* expression was four-fold increased when arsenate was added to the growth medium, whereas *TtsmtB* expression was two-fold increased in the presence of both arsenate and arsenite. The gene cloning and expression in *Escherichia coli*, followed by purification of the recombinant proteins, proved that, like ArsC of *S. aureus*, TtArsC was indeed a thioredoxin-coupled arsenate reductase and exhibited also weak phosphatase activity; TtSmtB was able to bind putative promoter regions of its own gene, of the operon including *TtsmtB* and of the operon including *TtarsC*; arsenite was antagonist of DNA binding by TtSmtB. The catalytic role of the first cysteine (Cys7) of TtArsC was also ascertained by site directed mutagenesis. All the results identify TtArsC and TtSmtB as the main actors in the arsenic resistance in *T. thermophilus* giving the first structural-functional characterization of thermophilic components of the arsenic detoxification mechanism.

Comprehensive knowledge on the molecular and genetic basis of detoxification, besides being stimulating from an evolutionary point of view, also represents an important starting point for developing efficient and selective environmental bioremediation approaches.

Chapter 1

GENERAL INTRODUCTION

1.1 Environmental pollutants: heavy metals and aromatic compounds

Increasing awareness of the hazards caused by environmental pollution has led to the search in many countries for methods to prevent the contamination of the environment and food. Among the main pollutants of natural origin there are hydrocarbons and heavy metals that, with the advent of industrialization, were mobilized from the ground and released in bio-available forms into the biosphere.

The heavy metals are normally present in the ecosystem, but human activities can contribute to their anomalous accumulation in the environment, favouring their insertion in the food chain and subsequently determining severe damages to all living organisms.

Among heavy metals and metalloids, the hazardous elements whose effects on humans have been fully characterized are Pb, Cd, Hg and As. Recently, arsenic has mainly attracted attention because of its toxicity and its ubiquity in the environment. It is an element that is widely distributed in the earth's crust, and it is naturally present in soil, water, air and all living matter from trace levels up to hundreds of $\text{mg}\cdot\text{kg}^{-1}$ (or $\text{mg}\cdot\text{l}^{-1}$) [80]. Arsenic exerts its toxicity by inactivating up to 200 enzymes, especially those involved in cellular energy pathways and DNA synthesis and repair. Acute arsenic poisoning is initially associated with nausea, vomiting, abdominal pain, and severe diarrhoea, followed by reported encephalopathy and peripheral neuropathy. Furthermore, arsenic is a well documented human carcinogen which affects numerous organs [86]. It has become evident that increasing human activities have modified the global cycle of arsenic, which is now ranked first in a list of 20 hazardous substances by the Agency for Toxic Substances and Disease Registry and United States Environmental Protection Agency [40]. It has been estimated that over 40 million people in the world are at risk from drinking arsenic-contaminated water [78]; for example, arsenic in the water supply in Bangladesh and West Bengal is considered to be a health catastrophe.

As with heavy metals and metalloids, aromatic compounds are also naturally present in the environment because they can derive from the degradation of polymers such as lignin. The 'natural source' of these molecules must not be misleading since, because of their extreme persistence in the environment, they represent a major cause of pollution. Major sources of such molecules are: (i) chemical and pharmaceutical industries that produce a wide array of xenobiotics and synthetic polymers; (ii) pulp and paper bleaching, which are the main sources of chlorinated organic compounds; (iii) fossil fuels (coal and petroleum), which may be accidentally released in large amounts into the ecosystem (oil spills) and whose combustion increases significantly CO_2 atmospheric levels (green-house effect) and causes deposition of nitric and sulfuric acids (acid rains and smog); and (iv) intensive agricultural activity, which releases massive amounts of fertilizers, pesticides, and herbicides [87].

Huge aromatic compound-contaminated sites exist worldwide; for example, in Germany more than 1400 coal tar sites contaminated by monoaromatic hydrocarbons, polycyclic aromatic hydrocarbons, phenolic and heterocyclic aromatic compounds were identified and only 20% of these areas has been considered for remediation [10]. These compounds exert their toxicity depending on their structure and chemical-physical properties; for example, nitro-aromatic compounds are poisonous by ingestion or subcutaneous, intraperitoneal and intramuscular routes; furthermore, they exhibit mutagenic and carcinogenic potential, decompose to emit

toxic fumes of NO_x, and are potent uncouplers of oxidative- and photo-phosphorylation [52].

From the above it is clear that the monitoring of water and soil pollution, mainly in proximity of industrial zones or contaminated areas, is one of the major environmental objectives. The removal of pollutants from the environment via natural physico-chemical and biological processes (natural attenuation) is, in general, a slow and unpredictable way of counteracting anthropogenic pollution and irreversible damage to the biosphere. Therefore, the main, if not the only, successful strategy to fight pollution is the use and manipulation of the detoxification abilities of organisms (bioremediation) [60]. Although most organisms are endowed with detoxification abilities (mineralization, transformation and/or immobilization of pollutants), bacteria and plants, particularly, have been the best-studied and the most frequently used for bioremediation strategies. In particular, microorganisms must have finely tuned mechanisms for sensing the level of pollutants in their growth environment and for controlling intracellular concentrations according to their biochemical needs. Generally, the mechanism underlying the detoxification ability of a pollutant involves its intra-or extracellular conversion, catalyzed by an enzyme, and/or its translocation to the outside of the cell, mediated by specific transport systems. Bacteria, which evolved more than three billion years ago, have developed strategies to obtain energy from virtually every compound; their abundance, together with their great ability of horizontal gene transfer and their high growth rates, allows them to evolve quickly and to adapt to changing environmental conditions, even to extreme environments that do not allow proliferation of other living beings.

In light of the above, microorganisms with such features may be a good starting point for the construction of biosensors to monitor environmental pollution or to contribute to the bioremediation processes.

1.2 Aromatic compound metabolism and detoxification

The interest in discovering how bacteria are dealing with hazardous environmental pollutants has stimulated important biochemical, genetic and physiological studies about the degradative abilities of microorganisms; this knowledge, in conjunction with genomic analysis, led to the creation of on line databases such as the Biocatalysis/ Biodegradation Database (<http://umbbdethz.ch/>) [101] or Kegg Database (<http://www.genome.jp/kegg/pathway.html>).

Microbial degradation of aromatic compounds has been extensively studied. By expressing different catabolic pathways, microorganisms can use a wide array of aromatic compounds as the sole carbon and energy sources. The aromatic catabolic pathways are usually composed of genes which are physically associated in operons and/or clusters encoding: enzymes involved in the aromatic catabolism, transporters responsible for active uptake of the compounds, and transcriptional regulators that modulate the expression of the genes encoding catabolic enzymes and transporters to the presence of the compounds to be degraded [28]. The catabolic clusters are often present in mobile genetic elements, such as transposons and plasmids, which facilitate their horizontal transfer and, therefore, rapid adaptation of microorganisms to the presence of new pollutants in a particular ecosystem. The main bacterial biochemical strategies to activate and cleave the aromatic ring depend primarily on the availability of oxygen [27]; however, many polluted environments are often anoxic, as aquifers, aquatic sediments and submerged soils; then biodegradation is carried out by either strict anaerobes or facultative microorganisms using alternative

electron acceptors, such as nitrate (denitrifying organisms), sulphate (sulfate reducers), Fe(III) (ferric-ion reducers), CO₂ (methanogens), or other acceptors (chlorate, Mn, Cr) [36].

In both aerobic and anaerobic biodegradation, structurally diverse compounds are degraded through many different peripheral pathways to a few intermediates that are further channelled to the Krebs cycle. In the classical aerobic catabolism, the hydroxylation and oxygenolytic cleavage of the aromatic ring are carried out by monooxygenases and/or hydroxylating dioxygenases and ring-cleavage dioxygenases, respectively. Most classical aerobic pathways converge to catecholic substrates which undergo either ortho- or meta- cleavage by intradiol or extradiol (type I and II) dioxygenases, respectively [27] (Fig. 1).

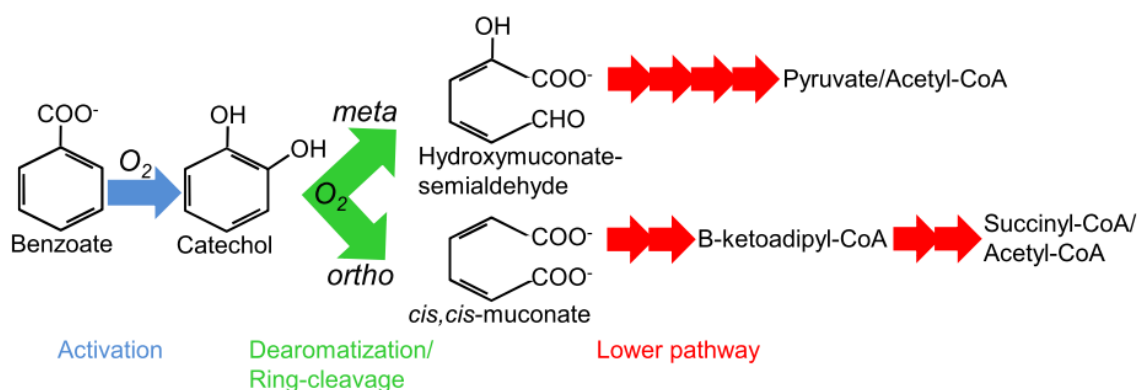


Fig. 1. Scheme of the classical aerobic biodegradation pathway to degrade benzoate in bacteria. An activation step (blue), a dearomatization/ring-cleavage step (green) and a degradation step to central metabolites (lower pathway) (red) can be identified.

In the anaerobic catabolism, the peripheral pathways converge to benzoyl-CoA, which becomes dearomatized by a specific multicomponent reductase that requires energy in the form of ATP [36] (Fig. 2).

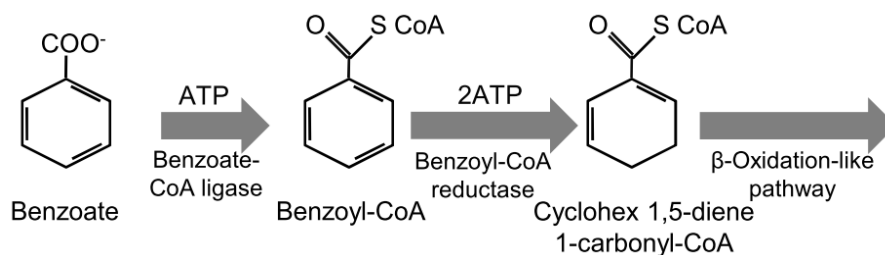


Fig. 2. Anaerobic degradation of benzoate.

Since many aromatic compounds are not only nutrients but also important chemical stressors for the bacteria, they constitute a good model system to study different aspects about their evolution/adaptation mechanisms [27]. Bacteria that dwell in polluted environments are often capable of evolving from pre-existing pathways that cope with natural compounds novel enzymes and regulators for the degradation of anthropogenic (xenobiotic) analogues, which have been in the biosphere for only a few years but whose toxic and mutagenic character impose a strong selective pressure [21]. Such microorganisms exhibit a wide variety of mono- and dioxygenases, which are useful for degradation of natural products and xenobiotics, efflux pumps that combat antimicrobial compounds of different origin, and a wide range of regulators that allow a rapid response to environmental changes [51]. Generally, such response can be initiated by the binding of transcription factors to

particular ligands, such as environmental signals; the bound complex has a different affinity for target regulatory sequences, resulting in a different access of the RNA polymerase to the promoters and hence a differential expression of one or more sets of genes [32].

Regulation of aromatic compound metabolism and detoxification

While metabolism is relatively well conserved in different organisms, regulation shows a wider diversity and, therefore, the whole understanding of the regulatory network of a given organism is a challenging task. The catabolism of aromatic compounds is controlled in a coordinated manner by complex regulatory networks that ensure that the balance between metabolic gain and stress endurance is not detrimental to the general cell physiology [27]. Specific transcriptional factors belonging to distinct families of regulators have been recruited and have evolved to control the expression of particular aromatic catabolic operons, thus ensuring the production of the enzymes and transporters at the right time. Some of these families include regulators involved in the aromatic compound catabolism, while, others are mainly involved in the stress response triggered by these compounds. The first class includes three families of widely characterized regulators: LTTRs, IclR- and AraC/XylS-type.

LysR-type transcriptional regulators (LTTR) comprise the largest family of prokaryotic proteins that regulate a target operon, such as CbnR, that controls the *cbnABCD* operon for chlorocatechol metabolism in *Ralstonia eutropha* [72], or NahR protein, that acts by controlling expression from both the *nah* operon, required for the metabolism of naphthalene to salicylate and pyruvate, and the *sal* operon, encoding the enzymes for salicylate conversion [94].

In general, the gene for a LTTR lies upstream of its target regulated gene cluster and is transcribed in the opposite direction. All identified LTTRs are small proteins (37 KDa) which act as transcriptional activators for their target metabolic operons in the presence of a chemical inducer, which is usually a pathway intermediate (e.g., salicylate, catechol, nitrotoluene). LTTRs repress their own expression, and both autorepression and activation of the catabolic operon promoter are exerted by the same binding site, which is called the regulator or repressor binding site (RBS) [11]. LTTRs interact with sites on the promoter DNA and recruit the α -CTD domain, thus increasing binding affinity of RNAP for the promoter and promoting transcription [101].

IclR-type regulators (25-30 kDa) are generally transcriptional repressors; but, those which control catabolic pathways have all been described as activators. Also in this case, the gene for the IclR-type regulator lies upstream of its target gene cluster and is transcribed in the opposite direction. The mechanism of autorepression is different among IclR-type regulators, since not all of them bind at the same position on the promoter region and since for some of them the addition of effectors changes the expression of the regulatory gene itself [101]. Probably, these regulators bind their promoter DNA in the absence of the effector to favor the recruitment of RNAP to the promoter, perhaps by optimizing the critical distance between the -35 and -10 elements [42].

AraC/XylS-type regulators for catabolic operons generally act as transcriptional activators in the presence of a chemical effector molecule. Most genes encoding for XylS-type regulators lie upstream of their target operon, but, in contrast to LysR- or IclR-type genes, they are transcribed in the same direction as the target genes. The

XylS/AraC-type regulators involved in the control of aromatic compounds degradation are typically dimers with a molecular mass of about 35 kDa, and contain two HTH motifs at their C-terminus.

Apart from the well-studied families described above, other members of regulatory families have been characterized (GntR, TetR and FNR). These proteins have common features: they are small proteins (around 27 kDa), which function as homodimers in solution, they have an HTH motif at their N-terminus and a dimerization domain including an effector binding pocket at their C-terminus. These regulators often act as transcriptional repressors in the absence of the substrates of the pathway, and the repression is released by the interaction with the aromatic compounds or one of their metabolites.

The most important family of transcriptional regulators involved in the response to environmental aromatic compounds is the MarR (Multiple Antibiotic Resistance Regulators) family. This family is constituted of ligand-responsive transcriptional regulators that are distributed throughout the bacterial and archaeal domains and includes proteins critical for the control of virulence factor production, the response to oxidative stress, the regulation of the catabolism of environmental aromatic compounds, the regulation of mechanisms of resistance to multiple antibiotics, organic solvents, household disinfectants, and pathogenic factors [4]. The genomic loci in which MarR homologs are encoded generally include divergent genes with the MarR homolog functioning as a regulator of both; usually, the binding site(s) reside in the intergenic region spanning those genes, but also distant genes may be regulated by a given homolog [83].

Identification of the MarR family began with the recognition of a chromosomally encoded mechanism of multidrug resistance in *Escherichia coli* K-12: the *marRAB* operon (Fig. 3) [18]. This operon encodes: i) a transcriptional repressor (MarR), which represses *marRAB* operon, ii) a transcriptional activator (MarA), which induces the expression of a number of genes responsible for the *mar* phenotype (including the AcrAB-TolC multidrug efflux system and *micF* that down-regulates the synthesis of the porin OmpF), iii) a small protein (MarB) of unknown function [1]. MarR binds to the intergenic region (*marO*) between *marC* gene (encoding a putative inner membrane protein of unknown function) and *marRAB* operon at two palindromic sequences with an apparent K_d of around 1 nM, obstructing the -10 and -35 promoter elements (Fig. 3) [61]. This mode of binding would sterically hinder the recruitment of RNA polymerase, hence repressing the transcription of both *marC* gene and *marRAB* operon [61].

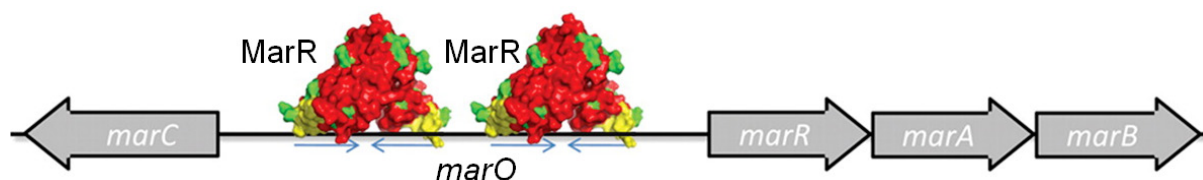


Fig. 3. Multiple Antibiotic Resistance (*mar*) locus from *E. coli*.

Additional mechanisms of transcriptional repression have been proposed for MarR homologs, including impeding promoter escape or transcriptional elongation by RNA polymerase and competing with a transcriptional activator for promoter binding [35]. Recently, MarR regulators acting as activators have been described. They bind to sites upstream of their target genes, where they may stabilize the RNA polymerase or compete with repressors for DNA binding [32].

Thus, it has been proposed that the choice of being an activator or a repressor could depend on the binding position relative to the promoter sequence. For example, *Streptomyces coelicolor* OhrR acts as a repressor as well as an activator by binding to the same operator region between *ohrR* and *ohrA* (encoding a protein involved in the organic hydroperoxides detoxification) [73]. The reduced form of OhrR binds cooperatively to the intergenic operator sequence covering the -10 and -35 promoter elements, thus hindering the recruitment of RNA polymerase; upon oxidation, OhrR binds loosely to *ohrA* promoting the RNA polymerase recruitment to the divergent *ohrR* gene, thus acting as an activator of *ohrR* transcription.

MarR-type regulators are relatively small proteins with a molecular mass between 17 and 22 kDa. These proteins have low sequence identity and share a triangular shape, they bind to their cognate palindromic or pseudopalindromic DNA (16-20 bp) as homodimers, resulting in either transcriptional repression or activation, or both [83]. Although sequence-specific in their binding, some MarR homologs are capable of associating with highly degenerate sequences [89]. MarRs may bind as a single dimer to an operator or as multiple dimers to adjacent sequences. The DNA binding domain of each monomer contains a variant of the common helix-turn-helix domain: the wHTH (winged helix turn helix) where the two wings, located at the corners of the triangle, are small beta sheets (Fig. 4A). The DNA-recognition helix makes sequence-specific contacts with the major groove of DNA, while the wings make different contacts, often with the minor groove or the backbone of DNA. Crystal structures of several MarR regulators have been obtained, either as apoproteins, in complex with the cognate DNA (Fig. 4B), or with various effectors, greatly contributing to the elucidation of the mechanistic basis of DNA and/or ligand binding.

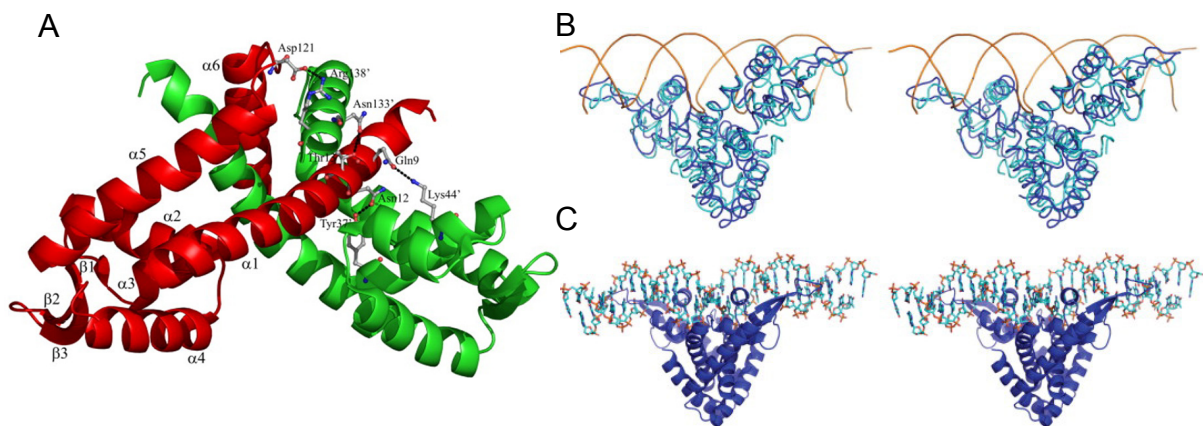


Fig. 4. (A) Ribbon representation of the structure of the BldR dimer of *S. solfataricus*, with one monomer shown in red and the other shown in green. Residues involved in polar interactions at the dimer interface are also shown. (B) A stereo view of the superposition of the OhrR-*ohrA* operator complex (PDB ID, 1Z9C) is shown. The OhrR dimer and the *ohrR* operator are colored cyan and orange, respectively. (C) A stereo view of the ST1710-DNA model is reported. The ST1710 is shown as a ribbon-model, and the DNA is shown as a stick model.

However, the identification of key residues involved in binding as well as those contributing to protein stability and/or dimerization has only been reported in a few cases. Structures generally comprise six α -helices and three β -strands (Fig. 4A) assuming an $\alpha 1 \alpha 2 \beta 1 \alpha 3 \alpha 4 \beta 2 \beta 3 \alpha 5 \alpha 6$ topology. The core of the domain consists of three α -helices ($\alpha 1$, $\alpha 2$, and $\alpha 3$) with $\alpha 2$ and $\alpha 3$ constituting the helix-turn-helix portion; the wing consists of two antiparallel beta sheets ($\beta 2$ and $\beta 3$), and is stabilized in part by a short beta strand ($\beta 1$). The N- and C- terminal helices

interdigitate to create a compact inter-subunit dimerization interface, which mainly consists of hydrophobic interactions and intermolecular hydrogen bonds (Fig. 5).

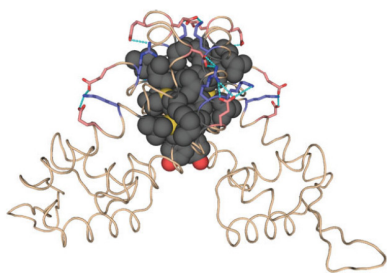


Fig. 5. Inter subunit interactions of the ST1710 dimer from *S. tokodaii*. Residues involved in hydrophobic interactions are shown in a black sphere model, and those involved in inter subunit salt bridge formation are shown in blue (positively charged residues) and red (negatively charged residues) stick models.

The dimerization domain (Fig. 5) dictates the distance between the DNA recognition helices, thus indirectly affecting the affinity of the protein for its cognate DNA; in fact, C-terminal deletions in MarR homologs have been shown to decrease their ability to form dimers, which correlates with attenuated DNA binding affinity [58]. The wing is vital for DNA interaction as evidenced by several mutational studies; for example substitutions of positively charged amino acids (R89, R90 and K91) in the wing abolish DNA binding of ST1710, a MarR member from *Sulfolobus tokodaii* [53]. The wing and recognition helices contact DNA directly or through water-mediated hydrogen bonding.

Another feature of MarR members is their ability to interact with specific ligands and, upon binding, to modulate DNA recognition; but in most cases the natural ligands are unknown. There are two main kinds of MarR effectors: many homologs bind small phenolic compounds, resulting in a conformational change that renders the proteins unable to bind DNA. The other variety of ligand interacts with its target to effect a transient covalent modification, which induces a conformational change, thus attenuating DNA binding; for example, this mechanism often occurs in oxidative stress responses and in the production of virulence factors [83].

E. coli MarR responds to a range of anionic lipophilic compounds such as 2,4-dinitrophenol, menadione and salicylate. The affinity of this regulator for its ligands is low; MarR binds salicylate with an apparent K_d of 0.5 mM [1]. Also DNA binding of EmrR is inhibited by structurally unrelated antibiotics, including the MarR ligands salicylate and 2,4-dinitrophenol [109]. EmrR is a MarR member that regulates the *emrRAB* operon in *E. coli*, which encodes EmrB, that pumps drugs across the cytoplasmic membrane, and EmrA that facilitates their passage through the periplasm.

The structure of *E. coli* MarR was solved with the proposed aromatic ligand salicylate (PDB ID: 1JGS) [2]. Two binding sites for salicylate (SAL-A and SAL-B) were identified in each subunit, each of which was relatively exposed on the surface and flanked the DNA recognition helix (Fig. 6A).

Subsequent literature suggested that the identified sites could be an artifact due to the high salicylate concentrations used to facilitate crystal packing; nevertheless the work of Alekshun *et al* led to a number of subsequent co-crystal structures of MarR homologs with salicylate: *Methanobacterium thermoautotrophicum* MTH313, *Salmonella typhimurium* SlyA, and *S. tokodaii* ST1710 (PDB IDs: 3BPX, 3DEU and 3GF2, respectively) [53,91], and *Staphylococcus epidermidis* TcaR [15]. Salicylate associates with MarR homologs and attenuates DNA binding only at high concentrations, questioning its biological relevance as a ligand. In MTH313 it is demonstrated that only the deeper binding site (SAL1 or SAL-A) is biologically relevant, as its occupancy imparts conformational changes in which the DNA

recognition helix is pushed away from its position in the apo structure and rotated of 5° . The structure of the archaeal homolog ST1710 was solved in its apo form and bound to either DNA or salicylate. Also ST1710 has low-affinity for salicylate (K_d is ≈ 20 mM) and the attenuation of DNA binding was only observed at 200–250 mM salicylate concentrations [53]. The ligand-bound structure is only modestly different from the apo structure, with distances between the recognition helices of 30° and only minor conformational changes in the wing. It has been proposed that salicylate is not the natural ligand for ST1710 and the hypothesis is that deprotonated salicylate may interact with MarR homologs non-specifically to equalize charges. *S. epidermidis*-encoded TcaR was crystallized in its native form and complexed with salicylate and with four antibiotics [15]: in contrast to the MTH313-salicylate structure, salicylate was found to occupy all four symmetry related sites in TcaR whereas the antibiotics interact with the protein at only two sites each. One site is near the DNA recognition helix, and it is highly solvent-exposed. Compared with the apo structure, salicylate- and antibiotic-bound structures show asymmetrical conformational alterations, primarily in the wHTH domain. The distances between the DNA recognition helices have not changed significantly compared with the apo structure. Results of all the structural analyses suggest that a common ligand-binding pocket is found in the crevice between the dimerization domain and the DNA-binding lobe (Fig. 6B).

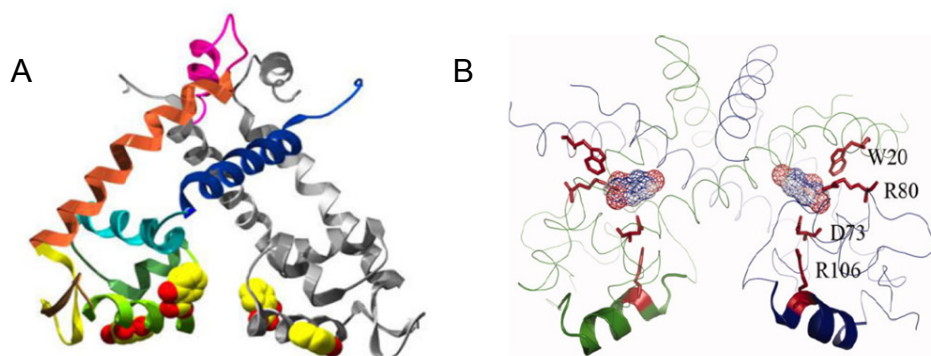


Fig. 6. (A) Ribbon diagram of MarR dimer with one subunit colored. Bound salicylate is shown in space-filling representation. (B) Structure of urate-docked HucR. The two monomers are in green and blue and the DNA binding helices are in cartoon representation. The residues identified to be important for ligand binding and attenuated binding to DNA are colored red. The ligand (urate) is in a mesh rendering.

A mechanism by which ligand-binding may induce a conformational change that propagates to the DNA recognition helix has been proposed for the binding of HucR to its ligand urate [84]. The mechanism by which urate-binding is communicated to the recognition helix works via a salt bridge between the Arg from the recognition helix and the Asp in $\alpha 2$; on binding of N3-deprotonated urate, a charge repulsion occurs that is propagated to the recognition helix. On the basis of this model, ligand binding in the deep binding pocket may lock the DNA-binding lobe by acting as 'molecular glue' between the dimerization domain and $\alpha 2$ of the wHTH domain, preventing conformational changes required for DNA binding. Salicylate binding to MTH313 and to *E. coli* MarR, urate binding to HucR and several other ligand interactions with MarR homologs show negative cooperativity [91,106], which may allow the MarR homolog to be responsive to a wider range of ligand concentration.

Some MarR homologs mediate responses to oxidative stress. The gene encoding OhrR from *Xanthomonas campestris* is co-transcribed with the adjacent gene, *ohr*, which is involved in protection against organic peroxides. OhrR is a repressor of the

gene *ohrA*, which encodes a thiol peroxidase that detoxifies organic hydroperoxides [50]. Oxidation-sensitive Cys residues in the C/N-terminal helices have been implicated in the mechanism of conformational change in order to attenuate DNA binding in OhrR [70]. In *Bacillus subtilis*, the *OhrR* and *ohr* genes are divergently encoded; *B. subtilis* OhrR represses transcription of *ohr* and although it is responsive to organic peroxides, its expression does not appear to be autoregulated [33]. *B. subtilis* OhrR contains a single reactive Cys residue per monomer whereas *X. campestris* OhrR contains two reactive Cys residues per monomer [54].

In summary, crystal structures reveal a conformational plasticity in MarR homologs that is exploited on binding of either cognate DNA or ligand; the spacing between recognition helices is essential for association with two consecutive DNA major grooves, and ligand-binding appears either to alter a DNA-compatible conformation into one that cannot interact with cognate DNA or to prevent conformational changes required for DNA binding in the case of homologs that do not come preconfigured for DNA binding. What is particularly notable is that such conformational changes originate in the region between the DNA binding lobe and the dimerization interface, whether due to binding of phenolic ligands or to cysteine modification.

Aromatic compound detoxification in Sulfolobus solfataricus P2

S. solfataricus P2 is an organism belonging to the domain Archaea. This is the third domain of life, comprehending microorganisms adapted to grow in extreme environments with regard to temperature, pH, ionic strength, and high concentrations of detergents and organic solvents. They are an evolutionary mosaic, being more similar to eukaryotes with respect to the macromolecular machinery and more similar to bacteria with respect to metabolic systems and genome organization [81]. For example, with regard to the transcription machinery, in most cases, homologues of bacterial regulators function in the context of the archaeal basal transcriptional apparatus, which resembles that of Eukarya [32,81]. The archaeal promoters are tripartite (Fig. 7); they have a INR element sequence near the transcription start site, an A+T rich sequence similar to the TATA box and a 6 bp element rich in purines (BRE-type) immediately upstream to the TATA box (Fig. 7). Transcription initiation is mediated by a single RNA polymerase, RNAP (containing between 8 and 13 subunits), and two general transcription factors: TATA element binding protein (TBP) and transcription factor B (TFB), a homologue of the transcription factor TFIIB, which binds to the BRE element and determines the directionality of the transcription (Fig. 7). This complex is sufficient to initiate transcription in cell-free systems, although an additional factor, transcription factor E, is required to increase transcription from some promoters [7]. Depending on the position, activation or inhibition of the different steps of the transcriptional initiation process has been established [99].

To date, little is known about the molecular mechanisms of transcriptional activation; it probably occurs by enhancing the recruitment of RNAP or basal transcription factors or by stabilizing their binding to the DNA target [77]. The negative modulation of transcription generally involves DNA binding proteins that bind or release target promoter DNAs in response to signaling ligands, and modulate transcription by competing for the binding sites of TBP/TFB or RNAP [77]. Archaea have attracted the attention of many protein chemists as models for understanding the molecular basis of adaptation to extreme conditions. Like all other living microorganisms, archaea are also able to defend against subtle changes to environmental conditions; in fact, their genomes encode finely regulated biochemical pathways for

detoxification as well as different regulative sequences designated to the response to stress agents.

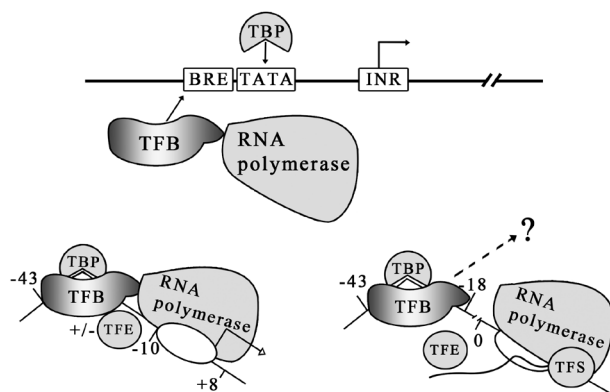


Fig. 7. An archaeal promoter region. TFB binding to the BRE is required to stabilize TBP binding to the TATA-box *in vitro*, and this occurs in the absence of RNA polymerase. However, as the figure suggests, there is evidence for TFB-RNAP complex formation in the absence of DNA and TBP. An archaeal preinitiation and postinitiation complexes are also showed.

In the archaeal domain transcriptional regulators belonging to MarR family have been widely studied; the crystal structures of four transcription factors: ST1710 from *S. tokodaii* [53], MTH313 from *M. thermoautotrophicum* [91], PH1061 from *Pyrococcus horikoshii* OT3 [74], and BldR from *S. solfataricus* [26] have been determined. The overall structure of all of these proteins is typical of the MarR family, particularly for elements taking part in the DNA-binding domain, while the most important differences are found in the dimerization domain. Comparison of the thermophilic ST1710 with mesophilic MarR showed an increase in the number of salt bridges on the protein surface predicted to be important in increasing the thermostability. Although the MarRs were characterized by structural or binding to ligands studies, in most cases their physiological role remains elusive.

Recently, the BldR transcriptional factor belonging to MarR family, involved in the aromatic aldehyde stress response, has been identified as part of a strategy for the adaptation of the archaeon *S. solfataricus* P2 to stress caused by aromatic aldehydes [32].

Some years later, it has been reported that *S. solfataricus* P2 is able to grow aerobically on phenol as the sole carbon source [46] and it has been hypothesised that *Sulfolobus* genus metabolize a large number of aromatic hydrocarbons such as cresols, benzene, toluene and polycyclic aromatic hydrocarbons. A genomic analysis revealed the existence of: a cluster of *orfs* coding for the subunits of a hypothetical bacterial multicomponent monooxygenase, an *orf* coding for a lower pathway protein of the catechol metabolism, and an *orf* coding for a putative catechol 2,3-dioxygenase [46]. In light of the above, the presence of aromatic catabolic enzymes and transcriptional regulators of the aromatic compound metabolism has been hypothesised in *S. solfataricus*.

In addition, Fiorentino *et al* reported that *S. solfataricus* has also a detoxification system that enables it to respond to aromatic compound stress [32]; in this study was demonstrated that the product of *Sso2536* (*adh*) gene, an alcohol dehydrogenase, is more efficient in the catalytic reduction of aldehydes than in the oxidation of the alcoholic counterparts and that *Sso2536* gene levels increase in cells treated with benzaldehyde [30]. BldR protein was identified and purified for its ability to bind specifically to the *Sso2536* regulatory sequences and its intracellular levels were demonstrated to increase upon exposure to the toxic benzaldehyde [32]. BldR is a MarR family member whose 3D structure has been solved at 1.9 Å [26]. It is the first example of a MarR member to be functionally characterized as a positive transcriptional regulator in Archaea [32]. Its gene is cotranscribed with an upstream

gene (*Sso1351*), coding for a multidrug efflux permease, and its expression increases in response to the addition of aromatic aldehydes, in a fashion identical to the trend observed for the *adh* gene.

A model has been proposed to explain the strategy of *S. solfataricus* to reduce the aromatic aldehyde concentration inside the cells. Increased binding of BldR to its own promoter upon benzaldehyde binding would induce auto-activation and increase the coexpressed drug export permease levels; the binding of BldR to *adh* promoter would also stimulate the gene transcription, the accumulation of the encoded enzyme, and hence the enzyme-catalyzed conversion of aldehydes to alcohols, which can be accumulated in the cell with minor damaging effect before being metabolized, via fission by aromatic ring cleavage enzymes, or extruded [32].

1.3 Arsenic detoxification strategies

Arsenic is a toxic metalloid belonging to the VA group of the periodic table. Its geochemical cycling is very complex; in fact, it involves chemical, physical and biological factors. Microbial activities play critical roles in the geochemical cycling of arsenic, because they can promote or inhibit its release from sediment material, mainly by redox reactions [76].

Arsenic is distributed in volatile forms (arsine) in the atmosphere, in mineral (arsenolite, arsenopyrite) and organic (dimethylarsinic acid, trimethylarsine, arsenocholine) forms in soils, and as inorganic dissolved compounds (arsenic acid, arsenious acid) in waters. This metalloid is present in various oxidation states: arsenate (+5), arsenite (+3), elemental arsenic (0) and arsenide (-3). The toxicity of arsenic depends on the type of compound, organic or inorganic compounds it forms, on its oxidation state, on how it is metabolized and accumulated. In general, inorganic forms are 100 times more toxic than organic compounds and the trivalent oxidation state (+3) is more toxic than the pentavalent (+5): $\text{AsH}_3 > \text{As(III)} > \text{As(V)} > \text{RAs-X}$.

As(V) toxicity is due to the fact that it can behave as a competitor of phosphate and can block the oxidative phosphorylation, decreasing the production of metabolic energy and leading to damage or death of the cell. Both in Archaea and Bacteria, As(V) can enter cells via two phosphate transport systems: phosphate specific transport (Pst) and Pit. Pit system uptakes phosphate and As(V) at similar rates, whereas the Pst is highly specific for phosphate [6]. Considering that As(V) is the thermodynamically favourable form of arsenic under aerobic conditions, more likely it is the most common form of arsenic in many environments and, in addition, it is less labile and toxic than As(III) [6].

As(III), instead, can cause the inactivation of some enzymes, by binding with high affinity to functional groups such as the thiolates of cysteine residues and the imidazolium nitrogens of histidine residues [88]; some studies have also shown that high concentrations of arsenite determine the inactivation of antioxidant enzymes (SOD) allowing the reactive oxygen species (ROS) to make serious damage to proteins, lipids and DNA. Differently from As(V), As(III), due to its un-ionized form at neutral pH, can passively move across the membrane bilayer or be transported by a carrier protein similar to those that transport un-ionized organic compounds such as glycerol and urea (aquaglyceroporin) [88].

The abundance and the toxicity of arsenic in the environment has guided the evolution of multiple resistance strategies in every organism studied, from *E. coli* to man. In fact, to counteract the deleterious effects of arsenic toxicity, the organisms

have evolved intriguing mechanisms which can include: reduced arsenic uptake by increased specificity of phosphate transporters [67]; transformation of the metalloid to less-toxic forms by methylation and adsorption [85,93]; arsenite oxidation (*aox* genes) [3]; respiratory arsenate reduction (*arr* genes); and arsenic resistance by arsenite extrusion (*ars* genes) (Fig. 8). Biological arsenate reduction occurs via two distinct pathways: one that allows to use arsenate as a terminal electron acceptor and a second that allows the cell to detoxify intracellular arsenate. For example, the arsenate respiratory pathway in *Shewanella* ANA-3 is composed of at least two proteins that are soluble in the periplasm: ArrA, a molybdenum-containing arsenate reductase, and ArrB, a 26 KDa Fe-S containing subunit. A membrane-associated c-type tetraheme cytochrome, CymA, is also required for arsenate respiration in several *arrAB*-containing *Shewanella* strains [68].

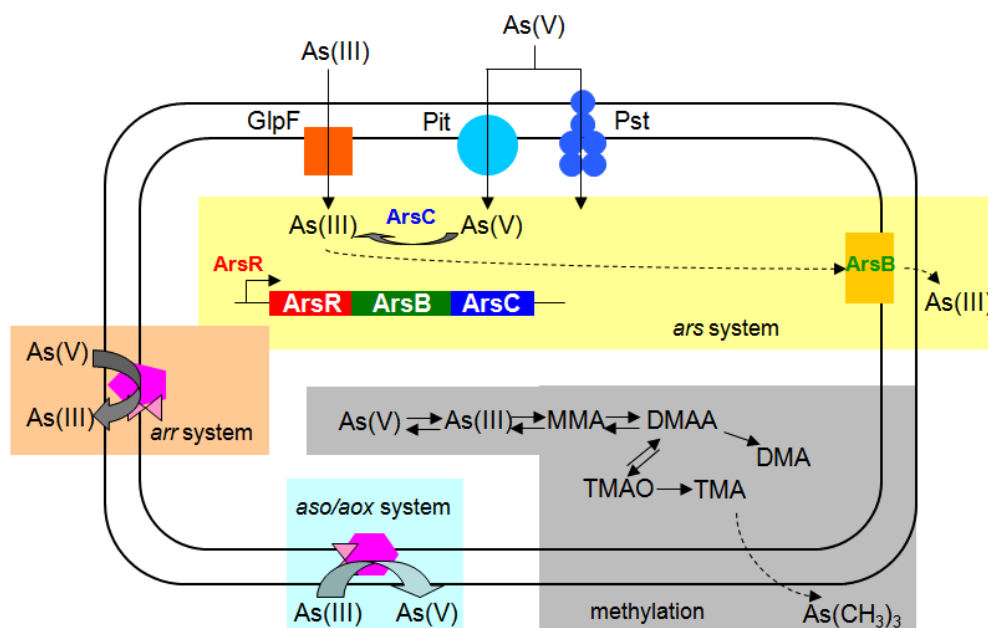


Fig. 8. Biochemistry of arsenic. Arsenic enters the cell through phosphate transporters or aquaglyceroporins. Arsenate is reduced to arsenite by *ArsC*, which extruded out of the cell by *ArsB* (*ars* system). Arsenite can serve as electron donor (*aso/aox* system) and arsenate as the ultimate electron acceptor during respiration (*arr* system). Inorganic arsenic can also be transformed into organic species (methylation). MMA (monomethyl arsenic acid), DMAA (dimethyl arsenic acid), DMA (dimethylarsinic acid), TMAO (trimethylarsine oxide), TMA (trimethyl arsine).

Ars systems (Arsenic Resistance System) allow the cell to detoxify intracellular arsenate and are widely distributed among bacteria and archaea. They have been well documented in *E. coli* [90], *Staphylococcus aureus* [48] and *B. subtilis* [92] and are controlled by genes encoding proteins that catalyze the two-electron reduction of As(V), which enters the cell as a phosphate analogue, followed by As(III) removal from the cell by a proton-driven arsenite transporter (Fig. 8) [80]. *Ars* genes can be carried on plasmids or chromosomes and are often organized in operons; the two most common contain either five (*arsRDABC*) or three (*arsRBC*) genes. The *arsRDABC* operon is found on the plasmids of Gram⁻ bacteria, such as *E. coli* R773, whereas *arsRBC* operon is found on the plasmids of Gram⁺ bacteria such as *S. aureus* pl258, or on bacterial chromosomes [80]. The *arsR* gene encodes a trans-acting repressor of the *ArsR*/*SmtB* family involved in transcriptional regulation, *arsB* encodes an arsenite efflux pump that exports arsenite but not arsenate, and *arsC* encodes a cytoplasmic arsenate reductase that converts arsenate to arsenite. Where

present, ArsA is an arsenite-stimulated ATPase, and ArsD is a metallochaperone that transfers trivalent metalloids to ArsA [57]. In addition to these well-studied components, a variety of *ars* clusters contains additional genes whose functions in arsenic resistance has not been clearly established. For example, *arsH* gene has been found in almost all Gram[−] bacteria that contain the *ars* operon but its function is not known yet [78]. A distinct gene (*arsM*), involved in the arsenic methylation, has been recently identified in more than 120 different archaea and bacteria [78], and characterized in *Rhodopseudomonas palustris* [85]. Methylation is thought to be a detoxification process even if not all the methylated products are less toxic than the inorganic forms. Methylated arsenicals can be generated by different processes, because the source of the methyl group and the transfer reaction can vary. In fungi, S-adenosyl methionine (SAM) serves as the donor, while anaerobic bacteria may use methyl cobalamin. ArsM, a methyltransferase, not only confers arsenic resistance, but also generates trimethylarsine gas. Arsenate conversion into the much more toxic arsenite before being transported out of the cytosol may seem contradictory. Why aren't there any arsenate-specific efflux systems? It can be speculated that, since the primordial atmosphere was not oxidizing, the majority of arsenic was in the form of As(III), and early organisms evolved detoxification mechanisms to cope with As(III). Once the atmosphere became oxidizing, probably the arsenite present in the oceans started being oxidized to arsenate so that the arsenate reductase evolution during the formation of an oxygenic world may have been rapid.

ArsC families

While the basic strategy to cope with arsenic is almost conserved among microorganisms, the arsenate reductases involved in detoxification are not. To date, three independently evolved families of arsenate reductases have been identified and characterized, which differ from each other in several of their physical and catalytic properties [64]. ArsC proteins are key enzymes in the arsenic detoxification process, since they reduce arsenate to arsenite, although using different strategies and reducing potential.

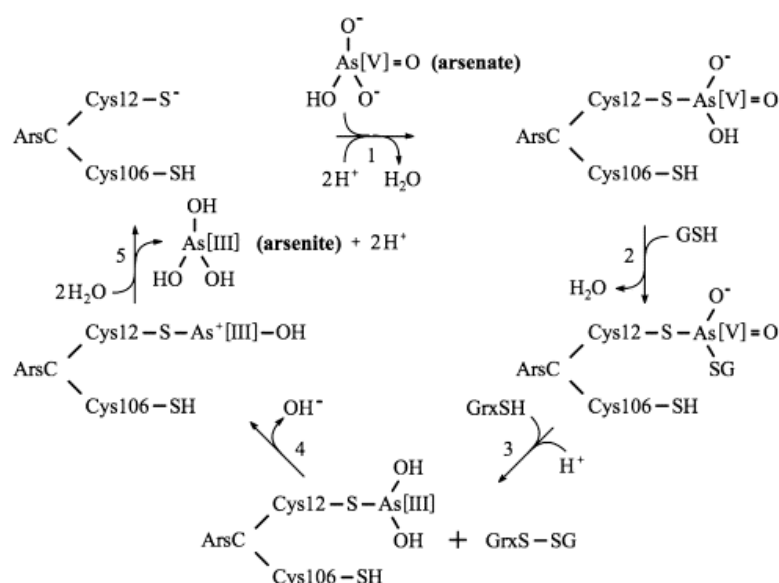


Fig. 9. Catalytic reaction of the Grx-coupled arsenate reductase of R773 plasmid.

The best-characterized family includes the first identified arsenate reductase of the *ars* operon of the *E. coli* R773 plasmid [16], but homologues have been found in many bacteria, both on plasmids and on chromosomes [88]. The catalytic mechanism of this family requires that arsenate binds to an anionic site consisting of three basic residues, Arg60, Arg94 and Arg107. In the next step, arsenate forms a covalent arsenate-thioester intermediate with the active site Cys12 (Fig. 9), which is then reduced in two steps by glutaredoxin and glutathione (Fig. 9). *E. coli* has three glutaredoxins, Grx1, Grx2 and Grx3, containing a Cys-Pro-Tyr-Cys consensus active site, which serves as a source of reducing potential for arsenate reduction. Grx2 is the preferred protein for such conversion [88].

The second family of arsenate reductases is related to proteins of the tyrosine phosphate phosphatase family. The characterized member is Acr2p from *Saccharomyces cerevisiae* [66]. Acr2p is related to the superfamily of phosphatases that includes Cdc25, a cell cycle phosphatase, and uses a HisCys(X)₅Arg motif in its active site [66]. Like the R773 ArsC, Acr2p has a single catalytic cysteine and uses glutaredoxin and glutathione as reductants. It does not exhibit phosphatase activity, even if an Acr2p mutant gained phosphatase activity and lost arsenate reductase activity. The ease by which a reductase can be changed into a phosphatase has led to propose that, under the selective pressure of ubiquitous arsenate in the environment, arsenate reductases have evolved from phosphatases.

The third family of arsenate reductases is typified by ArsC from *S. aureus* pl258 plasmid. This protein has very low sequence identity (10%) with the Gram[−] ArsCs (represented by *E. coli* ArsC) and distinct catalytic features, even if can catalyze the same chemical reaction and utilize a similar cysteine residue (Cys10 in *S. aureus* ArsC) at the active site. *S. aureus* ArsC has three redox active cysteines (Cys10, Cys82 and Cys89) critical for its enzymatic activity.

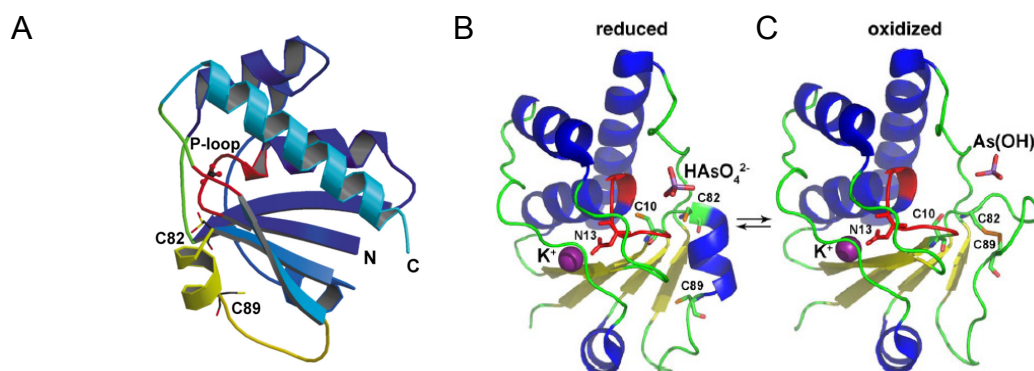


Fig. 10. Structure of pl258 ArsC. (A) Overall structure of the reduced form of arsenate reductase with the catalytic site shown in red, the portion involved in the redox function in yellow, the conserved Tyr binding site in LMW PTPases in green. (B) The structure of reduced pl258 ArsC (PDB 1LJL) and oxidized pl258 ArsC (PDB 1JFV) (C). α -helices are shown in blue; β -strands in yellow; and loops in green. Asn13 of the P-loop coordinates with potassium in the cation-binding site.

The first cysteine is included in a P-loop with the characteristic CX₅R sequence motif flanked by a β -strand and an α -helix (Fig. 10). This P-loop is a conserved motif of the LMWP (low molecular weight phosphatases) family. Cys10 is responsible for the initial nucleophilic attack on arsenate (Fig. 11); subsequently, Cys82 attacks the covalent intermediate and forms an intramolecular disulfide bridge. Cys89 completes the reaction forming a Cys82-Cys89 disulfide bond (Fig. 11), that can, in turn, be

reduced by a sequence of coupled redox reactions involving thioredoxin, thioredoxin reductase and NADPH [110].

A well characterized member of this family is ArsC from the Gram⁺ *B. subtilis*. The *arsC* gene of *B. subtilis* is located in the chromosome and is required to confer arsenate resistance *in vivo*. The corresponding protein conserves the three redox cysteines typical of the family and is a monomer in solution; the structure is a single $\alpha\beta$ domain containing a central four-stranded, parallel open-twisted β -sheet flanked by α -helices on both sides [8].

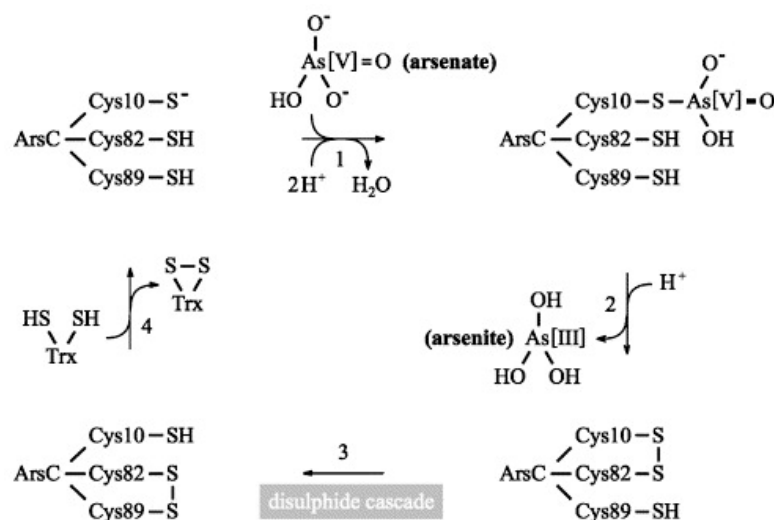


Fig. 11. Catalytic mechanism of p1258 ArsC.

The residues “CTGNSCR” in ArsC of *B. subtilis* form an oxyanion binding site named ‘arsenate binding loop’ (AB loop), which resembles the so called PTP loop (catalytic site of PTPases). The 3D structure of *B. subtilis* ArsC has revealed the importance of two further residues: Arg16, which acts by lowering the pKa values of the cysteines and stabilizing the thiolates needed during catalysis and Asp105, found in all Gram⁺ arsenate reductases, with a key role as general acid/general base catalyst [8]. The flexible loop and the three cysteines are essential for catalysis.

B. subtilis ArsC can hydrolyze p-nitrophenyl phosphate (pNPP), which is commonly used as the substrate in PTPase assays *in vitro*. p1258 ArsC also catalyzes the hydrolysis of pNPP, albeit 1000-fold slower than ‘acknowledged’ LMW PTPases. This dual catalytic activity (reductase and phosphatase activities) is limited to the Gram⁺ ArsC family, because Acr2p and R773 ArsC have no measurable phosphatase activities.

To date, the ArsC and Trx structures have been extensively investigated, and a study on the structure of the ArsC-Trx complex has also been reported [55]. Trxs are small proteins of 12 kDa, found in all organisms from bacteria to human, which catalyze the reduction of a wide range of downstream protein targets, one of which is the ArsC family. In the arsenate reduction pathway, during the Trx-ArsC interaction, a residue of Cys29 of Trx acts as the nucleophilic attacker and forms an intermolecular disulfide bridge with Cys89 of p1258 ArsC [55]. Extensive interactions between ArsC and Trx occur in the complex; anyway the solved structure of ArsC/Trx complex shows that the loop Cys82-Cys89 itself does not contribute much to the ArsC/Trx interaction.

In addition to the arsenate reductases described so far, other ArsCs have recently been identified, but their features did not allow to assign them to any of the above-mentioned classes. An example is represented by ArsC from *Synechocystis*

cyanobacteria, which has homology with the LMW-PTPases and *S. aureus* ArsC and, as the latter, exhibits reductase activity and a modest phosphatase activity [59]. Also *Synechocystis* ArsC has three catalytic cysteine residues (Cys8, Cys80 and Cys 82), but it uses glutathione and glutaredoxin as reducing potential. Its catalytic mechanism seems intermediate between *S. aureus* ArsC and *E. coli* and *S. cerevisiae* ArsCs: Cys8 operates the nucleophilic attack on the arsenate, and the other two cysteines are involved in the shuttle of the disulfide bridge on the enzyme surface, where it will be accessible to the reduction/regeneration.

In *Corynebacterium glutamicum* two atypical arsenate reductases have been characterized: ArsC1 and ArsC2 [75]. These enzymes exhibit a new catalytic mechanism that involves mycothiol (MSH), a pseudodisaccharide containing a reactive thiol, and mycothione reductase (MTR), a NADPH-dependent flavoenzyme. In the absence of ArsCs, MSH is capable of reducing As(V) to As(III), strongly suggesting the formation of a MS-As(V) adduct that will be reduced by MSH to generate As(III) and oxidized mycothiol, which will be further reduced by MTR with the consumption of NADPH. The active site thiolate in these ArsCs may work by lowering the energy barrier to facilitate adduct formation between arsenate and mycothiol. Arsenite might be released after a nucleophilic attack of MTR [75].

ArsR/SmtB family

The expression of heavy metal resistance genes is controlled at the transcriptional level by metal sensor proteins that 'sense' specific metal ions via their direct coordination [13]. The MerR and SmtB/ArsR families represent two general classes of transcriptional regulators that have endowed prokaryotes with the ability to respond to stress induced by heavy metal-toxicity. MerR-like proteins generally function as repressors in the absence of metal ions and become activators upon metal binding, by driving a metal induced DNA conformational switch that converts a sub-optimal promoter into a potent one [13].

The SmtB/ArsR members, perhaps distantly related to the MerR family, function exclusively as transcriptional repressors; when the apo-sensor proteins are bound to the operator/promoter (O/P) DNA, the resistance operons are repressed; metal binding strongly inhibits DNA binding [13].

One of the first characterized members of this family was SmtB from *Synechococcus* PCC 7942, which acts as a transcriptional repressor of *smtA* gene expression, encoding a class II metallothionein involved in sequestering excess Zn(II). Zn(II) is the preferred effector but SmtB also senses Co(II) and Cd(II). Other two Zn(II)-sensing regulators were recently identified: ZiaR from *Synechocystis* PCC 6803 and CzcA from *S. aureus*; the first regulates the *zia* operon, encoding ZiaR and ZiaA, a P-type ATPase metal efflux pump; the second negatively regulates the expression of the *czc* operon, which encodes, in addition to CzcA, a membrane-bound Zn(II) transporter of the cation diffusion facilitator family (CzcB) [13].

Besides the Zn(II)-sensors, others members of SmtB/ArsR family responsive to others metal ions exist: a Ni(II)/Co(II)-responsive regulator, NmtR, from *Mycobacterium tuberculosis* [82], which represses the *nmt* operon, containing the *nmtA* gene, whose protein product is a putative Co(II)-exporting ATPase transporter; and a plasmid- or chromosome- encoded ArsR that acts as the arsenic/antimony-responsive repressor of the *ars* operon in *E. coli* and other microorganisms.

X-ray crystallographic structure of dimeric apo-SmtB revealed that the protein is an elongated dimer with a two-fold axis of symmetry consisting of five α -helices and two

β -strands (Fig. 12) arranged into an $\alpha 1$ - $\alpha 2$ - $\alpha 3$ - $\alpha 4$ - $\beta 1$ - $\beta 2$ - $\alpha 5$ -fold. The primary subunit interface is formed between the two N-terminal $\alpha 1$ and C-terminal $\alpha 5$ helices [19]. The helix-turn-helix domain, particularly the sequence of the proposed DNA recognition α -helix ($\alpha 4$), is also highly conserved throughout the SmtB/ArsR family and is one of the distinguishing characteristics that define the membership.

Another distinguishing feature is the presence of an ELCV(C/G)D motif, which is required for metal ion sensing; this motif, termed 'metal binding box', is highly conserved among the members of the family. The X-ray structure of apo-SmtB revealed that the ELCV(C/G)D motif is located in the $\alpha 3$ helix, as part of the projected $\alpha 3$ -turn- $\alpha 4$ DNA binding motif [19]. Reports on ArsR have demonstrated that a substitution of one or both cysteines with non-metal binding residues in the ELCVCD motif inhibits the ability of arsenic to dissociate ArsR from the *ars* O/P. X-ray absorption spectroscopy of As(III)-ArsR complex revealed that As(III) is coordinated via three cysteine thiolate ligands within the putative $\alpha 3$ helix, two of which derive from the ELCVCD motif [95].

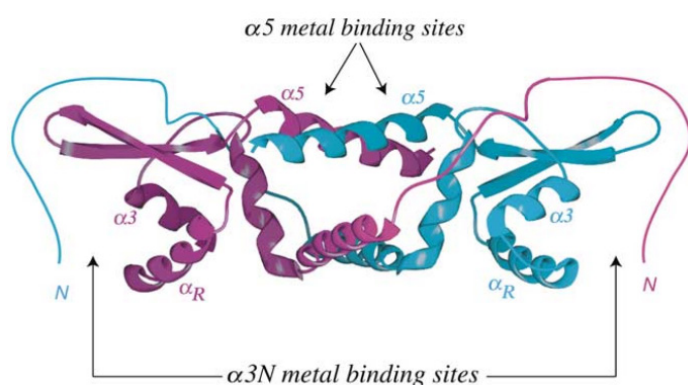


Fig. 12. A ribbon representation of the structure of homodimeric apo-SmtB determined crystallographically with one monomer shaded purple and the other blue.

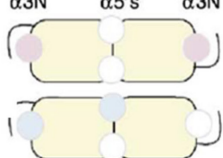
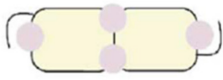
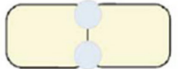
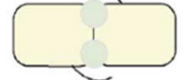
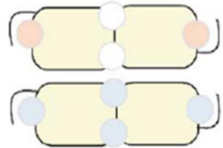

There is evidence that two distinct metal binding sites might exist in SmtB structure. In fact, mutagenesis of His105 and His106, located in the $\alpha 5$ helix of SmtB, have shown to inhibit Zn(II) sensing *in vivo*. This putative site in the $\alpha 5$ helix is not conserved in any SmtB/ArsR. In *Synechococcus* SmtB, *S. aureus* pl258 CadC, and *Synechocystis* ZiaR, two distinct metal binding sites have been identified, designated $\alpha 3$ N (or $\alpha 3$ in ArsR) and $\alpha 5$ (Fig. 13). The $\alpha 3$ N site is composed of three or four cysteines, two of which are located in the metal binding box and one or two additional cysteines derived from the N-terminal region [13]. Recent experiments suggest that the N-terminal thiolates are derived from one monomer and the $\alpha 3$ thiolates are derived from the opposite monomer to create an inter-subunit $\alpha 3$ N site [107]. ArsRs lack an N-terminal arm and coordinate the As(III) and Sb(III) ions with three clustered cysteines within the $\alpha 3$ helix, although just two of these are absolutely necessary for As(III) sensing *in vivo* [95].

Some SmtB/ArsR family regulators exclusively have the $\alpha 3$ / $\alpha 3$ N site (ArsR, CadC), others only contain the $\alpha 5$ / $\alpha 5$ C site (CzrA, NmtR) and others contain both sites (CadC, SmtB and ZiaR) (Fig. 13), probably because only one of them is essential for allosteric metalloregulation *in vitro* and metallo sensing *in vivo*.

SmtB/ArsR members bind to DNA promoter containing one or two imperfect 12-2-12 inverted repeats, generally in proximity or overlapping the transcriptional start site. The *smtA* promoter, for example, contains two imperfect inverted 12-2-12 repeats (S2/S1 and S4/S3), the first of which is required for full Zn(II) responsiveness of *smtA* expression, whereas the second has little, if any, effect on the regulation of *smtA* [102]. A site containing an imperfect 12-2-12 repeat located between the -10

sequence and the *czr* start site has also been identified in *czr* operator (Fig. 14) [96]. The stoichiometry of O/P binding by SmtB/ArsR repressors remains elusive. It was supposed that a single homodimer would bind to a single inverted repeat, with the DNA recognition helices ($\alpha 4$) interacting with successive major grooves. Studies with *Synechococcus* SmtB have revealed that the stoichiometry of binding is two tight-binding SmtB dimers to a 40-bp oligonucleotide containing a single S1/S2 inverted repeat. Interestingly, the full O/P region containing both S2/S1 and S4/S3 inverted repeats was found to just bind two dimers, rather than the four expected from the results obtained for each inverted repeat separately [13].

Fig. 13. Ligand binding sites of SmtB/ArsR family

Repressor	Metal Binding Sites	Metals	Regulatory Site
	$\alpha 3N$ $\alpha 5's$ $\alpha 3N$		
SmtB		Zn(II) Co(II)	$\alpha 5$
ZiaR		Zn(II)	$\alpha 5$ and $\alpha 3N?$
CzrA		Co(II) Zn(II)	$\alpha 5$
NmtR		Ni(II) Co(II)	$\alpha 5C$
CadC		Cd(II) Pb(II) Bi(III) Zn(II) Co(II)	$\alpha 3N$
ArsR		As(III) Sb(III)	$\alpha 3$

The stoichiometry of binding of ArsR-O/P has not been investigated, whereas footprinting experiments showed an ArsR-bound region of 33 bp overlapping the -35 element in the *ars* O/P [13].

The model which describes the mechanism of SmtB/ArsR repression is a derepression, in which metal binding by the sensor protein weakens the DNA binding affinity, so that RNA polymerase can load and initiate transcription of the operon. This was demonstrated by the reduction by 1000-fold of the binding affinity of SmtB dimers for a single 12-2-12 repeat upon coordination of a single equivalent of Zn(II), Cd(II) and Co(II) [104]. Also Pb(II)-, Cd(II)-, Zn(II)- and Bi(III)-bound CadC proteins have decreased affinities (340-, 230-, 185- and 170 fold respectively) for the DNA compared to the apo protein [12,14].

In a recent study an ArsR repressor from *Shewanella* able to regulate *arr* and *ars*

operons has been identified [68]. The arsenite-dependent expression of the *arr* operon shows regulatory characteristics similar to those of arsenic detoxification operons in other non arsenate-respiring prokaryotes such as *E. coli*, *Staphylococcus*, *Pseudomonas* and the archaeon *Halobacterium*.

<i>smt</i> S2/S1:	gcta AACACATGAACA	GT	T ATTCAGATATTcaaa
<i>smt</i> S4/S3:	ccacCATACCTGAATC	AA	GATTCAGATGTTaggc
<i>ziaA</i> :	cttt AA TATCTGAGCA	TA	TCTTCAGGTGTTtcaa
<i>czrAB</i> :	aatt AA TATATGAACA	AA	TATTCATATGAA agga
<i>nmtA</i> :	aataAA TAAATGAACA	TA	TGATCATATATT ctga
core motif:	aA tA xx TGAa ca	xx	ta tTCA xaTx t t
<i>cadCA</i> :	ataa TACTCAAAATA	AA	TATTTGAATGAAgatg
<i>arsRBC</i> :	tggt ATTAAT CATATG	CG	TTTTTGGTTATG tgt t
core motif:	TAxAx T CAAAt a	xx	ta TTT Ga xTxTA

Fig. 14. Alignment of the imperfect 12-2-12 inverted repeats from the *Synechococcus* *smt* O/P (S2/S1, S4/S3), *Synechocystis* *zia* O/P, *S. aureus* *czr* O/P, *M. tuberculosis* *nmt* O/P, *S. aureus* *pl258 cad* O/P, and *E. coli* R773 *ars* O/P.

To date, the common features of the characterized ArsR proteins include their predicted size (12 to 16 kDa), helix-turn-helix domain and several conserved cysteine

residues. ArsR represses transcription in the absence of arsenite by binding near the *ars* promoter region, the repression is usually alleviated in the presence of arsenite and antimonite but not arsenate. Several cysteines are known to mediate the interactions with trivalent arsenic or antimony oxyanions, for example, in *E. coli* ArsR mutations in the Cys32 and Cys34 residues render the protein insensitive to inducers while preserving the protein's ability to bind DNA, causing *ars* expression repression when arsenite is present [68].

Arsenite extrusion: ArsB and ArsA

In bacteria there are two basic mechanisms of arsenite extrusion: one is an arsenite carrier-mediated efflux, where energy is supplied by the membrane potential of the cell, and the other is an arsenite-translocating ATPase [25]. The majority of bacteria use ArsB, found in most *ars* operons, as an efflux pump to extrude arsenite. When ArsA is co-expressed with ArsB, an ArsAB complex is formed that is obligatorily coupled to ATP.

ArsB is a 429-residue integral membrane protein with 12 membrane-spanning segments [108] (Fig. 15). It appears to be a trivalent metalloid/ H^+ antiporter that uses the membrane potential, positive at the exterior, to extrude arsenite. When ArsA is synthesized in *E. coli* containing the *arsRDABC* operon, cells are more resistant to arsenite because the ArsAB ATPase is much more efficient than ArsB alone [63]. The 583-residue ArsA ATPase is a member of a family of ATPases that probably arose from GTPases. It is normally bound to ArsB [24] but, in the absence of ArsB, ArsA is found in the cytosol and can be purified as a soluble protein. The crystal structure of ArsA has been determined [111]. It consists of three domains: two nucleotide-binding domains (NBD) and a single arsenite-binding domain (MBD). MBD is an allosteric site and binds three Sb(III) or As(III) ions, activating the ArsA ATPase activity. As(III)/Sb(III) stimulate ATP hydrolysis by binding three cysteines located in different places in the primary sequence, so the protein must fold to bring the three residues close enough together to interact with the metalloid [88].

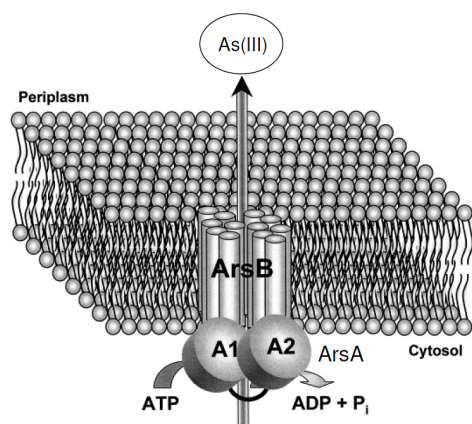


Fig. 15. The ArsAB pump. The complex of ArsA and ArsB forms an anion-translocating ATPase that catalyzes extrusion of arsenite or antimonite. ArsA has two homologous halves, A1 (N-terminal) and A2 (C-terminal).

This dual mode of energy coupling (ArsB alone or ArsAB complex) has led to propose that the ArsAB pump has evolved by the association of a secondary carrier with a soluble ATPase.

Arsenic metabolism in Thermus genus

Hot spring waters typically contain 1-10 mg/l As, but can reach up to 50 mg/l [5]. Even though arsenic is a common trace constituent of geothermal fluids, little work

has been conducted to study the microbial processes that may impact arsenic cycling in these settings. The natural discharge of these fluids often results in increased arsenic levels in surface waters and aquifers surrounding hot springs. For example, at Yellowstone National Park, over 100000 kg of geothermally derived arsenic is estimated to leave the western boundary each year, affecting water quality within a large region [71].

Biochemists and microbial ecologists have been investigating hot springs to explore the diversity and physiology of extremophiles and to seek biomolecules with potential technological importance. The exploration of life in extreme environments and the analysis of As-contaminated sites have led to the isolation of a large number of hyperthermophiles which are able to reduce arsenate or oxidize the arsenite [9,105] for their metabolism. Starting from these studies, it has been hypothesized that arsenic resistance could have evolved in geothermal environments where microorganisms have been exposed to natural sources of arsenic. Inspection of the genomes of different thermophilic bacteria and archaea has revealed putative genes involved in arsenic detoxification. For example, in the *Thermus* genus, *T. aquaticus* and *T. thermophilus* have been found to rapidly oxidize arsenite [38] and it has been shown the ability of *Thermus* HR13 to use arsenate for respiration [37]. *T. aquaticus* and *T. thermophilus* are not able to grow with arsenite as a sole energy source and catabolic energy does not derive from arsenite oxidation. Considering that arsenate is less toxic than arsenite, the ecological role of arsenite oxidation by these organisms is likely detoxification [37].

Moreover, *Thermus* HR13 has the capacity to reduce As(V) in addition to the potential for As(III) oxidation, probably as a respiratory process during periods of anoxia [37]. Studies have also shown the isolated *Thermus* sp. SA-01 to be a facultative anaerobe capable of using arsenate, nitrate, ferric iron and elemental sulphur as terminal electron acceptors [49].

The *Thermus* genus is known to be very widespread, inhabiting nearly all types of thermally influenced circumneutral waters that have been examined, including terrestrial hot springs as well as domestic and industrial sources. *T. thermophilus*, in particular, is a Gram⁻ negative aerobic bacterium which can grow at temperatures ranging from 47°C to 85°C, originally isolated from a natural thermal environment in Japan. The strain exhibits high competence for natural transformation and, therefore, is amenable to genetic manipulation [41]. Furthermore, the genome of *T. thermophilus* HB27 has been sequenced [43].

1.4 Biomonitoring and bioremediation

Traditionally, chemical methods are used for the detection of pollutants; recently, they have been complemented by biological techniques. For example, the environmental heavy metal pollution is determined by quantification of total metals by treatment with acids to solubilise the metal ion from solid matrix, and then evaluated with conventional analytical methods, such as atomic absorption spectrometry and ion chromatography. In order to detect aromatic compounds in aqueous solution, a combination of solid-phase micro-extraction (SPME) and infrared (IR) attenuated total reflection spectroscopic methods is used. These techniques are expensive, time-consuming and often need substantial sample pre-treatment. The biosensor, a recent product of biotechnology, has attracted considerable attention as a possible successor to a wide range of analytical techniques. The use of biosensors as environmental monitoring system has many advantages, including specificity, speed,

portability, ease of use, real-time signalling and, moreover, it allows toxicological measurements, whereas conventional techniques can only provide concentration values.

To construct a whole-cell biosensor, the recombinant strain must be characterized by a genetic system comprising an inducible promoter, responsive to toxic substances, fused to a reporter gene encoding an easily detectable protein. A variety of well-characterized promoters are available for genetic manipulations, for example promoters of various heavy metals [34,44], aromatic aldehydes [31], hydrocarbons [79], pesticides [17], and salicylates [45]. The most-used reporter genes are: *lacS*, encoding a β -galactosidase, whose enzymatic activity can be determined by a colorimetric reaction; luciferase genes (*lux*) [65] or the gene for Green Fluorescent Protein (GFP) [56], that can be detected by bioluminescence and spectrofluorimetry, respectively.

Many examples of biotechnological applications of engineered microorganisms for environmental monitoring of heavy metals are reported; for example, a recombinant bacterial biosensor that can be employed as a primary screening technique for aquatic heavy metal pollution has been constructed [39]: *zntR* gene-*zntAO/P-gfp* gene cassette has been introduced on a plasmid vector into *E. coli* DH5 α and simple fluorescence measurements are well correlated to metal concentrations. *ZntR* and *zntA*, the Zn(II) responsive elements were used as receptors elements controlling the expression of *gfp* reporter gene. This biosensor is able to respond to toxic levels of Zn(II), Cd(II) and Hg(II) with a maximum recorded fluorescence at 20 ppm, 0.005 ppm and 0.002 ppm, respectively [39].

A cell-whole biosensor to detect aqueous concentrations of aromatic aldehydes has also been reported [31]. The biosensor was based on an *E. coli* B121DE3 strain containing: a hybrid transcriptional fusion between an archaeal promoter responsive to aromatic aldehydes (*S. solfataricus adh* promoter) and the *gfp* gene, and the gene for the sensor protein BldR [32], carried on a different compatible plasmid. This system responded to millimolar concentrations of benzaldehyde, cinnamaldehyde and salicylaldehyde [31].

Together with cellular biosensors to detect heavy metals, several examples of enzyme biosensors have also been reported in literature, such as the use of the inhibitory effect of mercury on the urease enzyme to obtain indirect concentration measurements of the metal [69]. Enzyme-based biosensors have high specificity and selectivity, but their use is limited by the low half-life, the high purification costs and the need for cofactors/coenzymes.

Another possibility to use the detoxification strategies of microorganisms for biotechnological applications is the bioremediation. Ideally, bioremediation strategies should be designed based on knowledge of the microorganisms which inhabit the polluted environments, their metabolic abilities, and how they respond to changes in environmental conditions. To enhance the metabolic efficiency of a microorganism for a particular environmental application, genetic engineering should be carried out at two different levels: manipulation of the specific catabolic pathway, and manipulation of the host cell. In order to improve the rate of pollutant removal and broaden the range of substrates of a catabolic pathway, manipulation of the key enzymes and the regulatory mechanisms that control the expression of the catabolic genes is required [22]. Protein stability and activity can be altered and/or improved by protein engineering and directed molecular evolution techniques (DNA shuffling and other *in vitro* recombination methods). Moreover, metabolic engineering allows the generation of novel hybrid pathways by assembling catabolic modules from different

origins in the same host cell, thus leading to pathway expansion to new substrates, completion of incomplete pathways, the creation of new routes, and construction of bacteria with multiple pathways [87]. The rational combination of catabolic pathways may allow the complete metabolism of xenobiotics, as has been shown with the development of bacteria capable of mineralizing PCBs.

Substances such as heavy metals cannot be destroyed or biodegraded. Recombinant microorganisms have been developed to accumulate and/or immobilize heavy metals present in soil and water [62,103]. As some metals can serve as terminal electron acceptors in microbial respiration, anaerobes that use them have been applied to reduce the soluble oxidized form to the insoluble form, thereby preventing its further spread in the environment. Sinha *et al* have reported a study on the use of *Enterobacter* sp. EMB21 for mercury remediation wherein the metal can be trapped inside the cells using the intrinsic resistance system of the microorganism [97]. Previously, Deng and Jia developed a recombinant photosynthetic bacterium *R. palustris*, expressing mercury transport system and metallothionein, simultaneously, for Hg(II) accumulation [23].

Biosafety is a major issue when releasing recombinant microorganisms into any open environment. Several genetic circuits have been developed to allow survival of the recombinant microorganisms only when present in the polluted site and during the time required for removal of the pollutant (biological containment). To avoid dispersal of the recombinant trait to the native microbial population, different gene-containment circuits based on a toxin and its cognate antidote have also been developed. Such active containment systems significantly reduce the potential risks that release of recombinant bacteria might cause to the ecosystem [100].

1.5 The aim of the work

Although microorganisms have acquired the ability to use pollutants as carbon and energy sources, their efficiency in removing such pollutants might not be optimal for the elimination of present-day pollution. In fact, microorganisms have evolved towards their ecological fitness rather than biotechnological efficiency; thus, it would take a long natural selection-driven time for bacteria to evolve their means for cleaning an environment from anthropogenic pollution.

Hence, the study of the physiology, biochemistry and genetics of the catabolic pathways to recreate and accelerate natural processes for different biotechnological applications becomes crucial.

To date, many mesophilic microorganisms have been employed for bioremediation or biomonitoring. The advantages of using thermophilic microorganisms can lie in their ability to survive to harsh conditions (resistance to caotropic agents or detergents often present in industrial off-loads and wastewaters); thus, the thermophilic microorganisms could be good candidates for the construction of more stable and stronger cellular or enzymatic biosensors.

For this purpose a thorough characterization of the molecular mechanisms of response to environmental stress as well as the identification and characterization of active and stable enzymes under extreme conditions is required.

The goals of this research project are the molecular characterization of the detoxification pathways from aromatic compounds and arsenic in thermophilic microorganisms, in order to obtain a better understanding of the mechanisms underlying the resistance to such toxic compounds.

The first step of this research has been the physiological study of *S. solfataricus* by *in silico* identification and molecular characterization of the detoxification pathway from aromatic compounds (benzaldehyde and salicylate) and its regulation. The aromatic compound detoxification pathway in the crenarchaeon *S. solfataricus* has already been characterized, and revealed the presence of a regulator/efflux pump system in this organism [32]. In fact, the identification of regulatory sequences responsive to aromatic compounds and of associated transcriptional regulators (Lrs14 and BldR) helped formulate a model in which the detoxification of these substances is obtained by the use of a membrane permease and of an alcohol dehydrogenase, which are under the control of the aforementioned regulatory systems [30,32]. Little is currently known about the interactions between the regulatory proteins and the effector compounds and it is not clear how effector binding to the regulatory proteins transmits the “activation” signal for RNAP. Better knowledge on the binding affinities to the effectors and the role of the ternary complex (protein, effector, target DNA) may help in the development of bacterial regulatory systems to be used as environmental quality sensors, which should have sensitivities below micromolar concentrations.

The second step of this research has been the physiological study of *T. thermophilus* by *in silico* identification and characterization of the arsenic detoxification pathway and its regulatory circuits. In most organisms arsenic resistance has only been characterized in terms of their ability to grow in the presence of high arsenic concentrations (over 250 mM of arsenate) [47]; to date, among thermophiles, *ars* systems and their genetic determinants have been preliminarily characterized in *Geobacillus kaustophilus* and *Acidithiobacillus caldus* only [20,29]. With the aim of identifying thermophilic proteins/enzymes involved in metalloid detoxification, a preliminary analysis of the genome of *T. thermophilus* was carried out. Results showed the presence of an *orf* encoding a putative transcriptional repressor belonging to SmtB/ArsR family, TtSmtB, located in the genome upstream of another *orf* which encodes a putative cation transporting ATPase containing the HMA (Heavy-Metal-Associated) motif; such motif is composed of 30 amino acid residues and has been found in proteins involved in transport and detoxification of metals. In addition, *T. thermophilus* also contains in its genome the gene encoding a putative arsenate reductase (ArsC) and the gene encoding a putative efflux pump for the toxic arsenite.

The above suggests the possibility of building biological systems (cellular or enzymatic) for the traceability or remediation of pollutants after a thorough molecular, structural and functional characterization of the components involved and their interactions.

References

- [1] Alekshun MM, Levy SB (1999). The *mar* regulon: multiple resistance to antibiotics and other toxic chemicals. Trends Microbiol; 7:410-413.
- [2] Alekshun MN, Levy SB, Mealy TR, Seaton BA, Head JF (2001). The crystal structure of MarR, a regulator of multiple antibiotic resistance, at 2.3 Å resolution. Nat Struct Biol; 8:710-714.
- [3] Andreoni V, Zanchi R, Cavalca L, Corsini A, Romagnoli C, Canzi E (2012). Arsenite oxidation in *Ancylobacter dichloromethanicus* As3-1b strain: detection of genes involved in arsenite oxidation and CO₂ fixation. Curr Microbiol; 65:212-218.
- [4] Aravind L, Anantharaman V, Balaji S, Mohan Babu M, Iyer LM (2005). The many faces of the helix-turn-helix domain: transcription regulation and beyond. FEMS Microbiol Rev; 29:231-262.
- [5] Ballantyne JM; Moore JN (1988). Arsenic geochemistry in geothermal systems. Geochim Cosmochim Acta; 52:475-483.

- [6] Bartolucci S, Contursi P, Fiorentino G, Limauro D, Pedone E (2013). Responding to toxic compounds: a genomic and functional overview of Archaea. *Frontiers in Bioscience*; 18:165-189.
- [7] Bell SD, Brinkman AB, van der Oost J, Jackson SP (2001). The archaeal TFII α homologue facilitates transcription initiation by enhancing TATA-box recognition. *EMBO Rep*; 2:133-138.
- [8] Bennett MS, Guan Z, Laurberg M, Su XD (2001). *Bacillus subtilis* arsenate reductase is structurally and functionally similar to low molecular weight protein tyrosine phosphatases, *Proc Natl Acad Sci USA*; 98:13577-13582.
- [9] Blum JS, Kulp TR, Han S, Lanoil B, Saltikov CW, Stolz JF, Miller LG, Oremland RS (2012). *Desulfohalophilus alkaliarsenatis* gen. nov., sp. nov., an extremely halophilic sulfate- and arsenate- respiring bacterium from Searles Lake, California. *Extremophiles*; 16:727-742.
- [10] Blum P, Sagner A, Tiehm A, Martus P, Wendel T, Grathwohl P (2011). Importance of heterocyclic aromatic compounds in monitored natural attenuation for coal tar contaminated aquifers: A review. *J Contam Hydrol*; 126:181-194.
- [11] Bundy BM, Collier LS, Hoover TR, Neidle EL (2002). Synergistic transcriptional activation by one regulatory protein in response to two metabolites. *Proc Natl Acad Sci USA*; 99:7693-7698.
- [12] Busenlehner LS, Apuy JL, Giedroc DP (2002). Characterization of a metalloregulatory bismuth(III) site in *Staphylococcus aureus* pl258 CadC repressor. *J Biol Inorg Chem*; 7:551-559.
- [13] Busenlehner LS, Pennella MA, Giedroc DP (2003). The SmtB/ArsR family of metalloregulatory transcriptional repressors: structural insights into prokaryotic metal resistance. *FEMS Microbiol Rev*; 27:131-143.
- [14] Busenlehner LS, Weng TC, Penner-Hahn JE, Giedroc DP (2002). Elucidation of primary (K3N) and vestigial (K5) heavy metal binding sites in *Staphylococcus aureus* pl258 CadC: evolutionary implications for metal ion selectivity of ArsR/SmtB metal sensor proteins. *J Mol Biol*; 319:685-701.
- [15] Chang YM, Jeng WY, Ko TP, Yeh YJ, Chen CK, Wang AH (2010). Structural study of TcaR and its complexes with multiple antibiotics from *Staphylococcus epidermidis*. *Proc Natl Acad Sci USA*; 107:8617-8622.
- [16] Chen CM, Misra TK, Silver S, Rosen BP (1986). Nucleotide sequence of the structural genes for an anion pump. The plasmid-encoded arsenical resistance operon. *J Biol Chem*; 261:15030-15038.
- [17] Chouteau C, Dzyadevych S, Durrieu C, Chovelon JM (2005). A bi-enzymatic whole cell conductometric biosensor for heavy metal ions and pesticides detection in water samples. *Biosens Bioelectron*; 21:273-281.
- [18] Cohen SP, Hachler H, Levy SB (1993). Genetic and functional analysis of the multiple antibiotic resistance (*mar*) locus in *Escherichia coli*. *J Bacteriol*; 175:1484-1492.
- [19] Cook WJ, Kar SR, Taylor KB, Hall LM (1998). Crystal structure of the cyanobacterial metallothionein repressor SmtB: A model for metalloregulatory proteins. *J Mol Biol*; 275:337-346.
- [20] Cuebas M, Sannino D, Bini E (2011). Isolation and characterization of arsenic resistant *Geobacillus kaustophilus* strain from geothermal soils. *J Basic Microbiol*; 51:364-371.
- [21] de Las Heras A, Chavarri'a M, de Lorenzo V (2011). Association of *dnt* genes of *Burkholderia* sp. DNT with the substrate-blind regulator DntR draws the evolutionary itinerary of 2,4-dinitrotoluene biodegradation. *Mol Microbiol*; 82:287-299.
- [22] De Lorenzo V (2001). Cleaning up behind us. *EMBO Rep*; 2:357-359.
- [23] Deng X, Jia P (2011). Construction and characterization of a photosynthetic bacterium genetically engineered for Hg²⁺ uptake. *Bioresour Technol*; 102(3):3083-3088.
- [24] Dey S, Dou D, Tisa LS, Rosen BP (1994). Interaction of the catalytic and the membrane subunits of an oxyanion-translocating ATPase. *Arch Biochem Biophys*; 311:418-424.
- [25] Dey S, Rosen BP (1995). Dual mode of energy coupling by the oxyanion-translocating ArsB protein. *J Bacteriol*; 177:385-389.
- [26] Di Fiore A, Fiorentino G, Vitale RM, Ronca R, Amodeo P, Pedone C, Bartolucci S, De Simone G (2009). Structural analysis of BldR from *Sulfolobus solfataricus* provides insights into the molecular basis of transcriptional activation in archaea by MarR family proteins. *J Mol Biol*; 388:559-569.
- [27] Díaz E, Jiménez JI, Nogales J (2012). Aerobic degradation of aromatic compounds. *Curr Opin Biotechnol*; 24:1-12.
- [28] Diaz E, Prieto MA (2000). Bacterial promoters triggering biodegradation of aromatic pollutants. *Curr Opin Biotech*; 11:467-475.

- [29] Dopson M, Lindström EB, Hallberg KB (2001). Chromosomally encoded arsenical resistance of the moderately thermophilic acidophile *Acidithiobacillus caldus*. *Extremophiles*; 5:247-255.
- [30] Fiorentino G, Cannio R, Rossi M, Bartolucci S (2003). Transcriptional regulation of the gene encoding an alcohol dehydrogenase in the archaeon *Sulfolobus solfataricus* involves multiple factors and control elements. *J Bacteriol*; 185(13):3926-3934.
- [31] Fiorentino G, Ronca R, Bartolucci S (2009). A novel *E. coli* biosensor for detecting aromatic aldehydes based on a responsive inducible archaeal promoter fused to the green fluorescent protein. *Appl Microbiol Biotechnol*; 82:67-77.
- [32] Fiorentino G, Ronca R, Cannio R, Rossi M, Bartolucci S (2007). MarR-like transcriptional regulator involved in detoxification of aromatic compounds in *Sulfolobus solfataricus*. *J Bacteriol*; 189(20):7351-7360.
- [33] Fuangthong M, Atichartpongkul S, Mongkolsuk S, Helmann JD (2001). OhrR is a repressor of *ohrA*, a key organic hydroperoxide resistance determinant in *Bacillus subtilis*. *J Bacteriol*; 183:4134-4141.
- [34] Fujimoto H, Wakabayashi M, Yamashiro H, Maeda I, Isoda K, Kondoh M, Kawase M, Miyasaka H, Yagi K (2006). Whole-cell arsenite biosensor using photosynthetic bacterium *Rhodovulum sulfidophilum*. *Rhodovulum sulfidophilum* as an arsenite biosensor. *Appl Microbiol Biotechnol*; 73:332-338.
- [35] Gala'n B, Kolb A, Sanz JM, Garcí'a JL, Prieto MA (2003). Molecular determinants of the *hpa* regulatory system of *Escherichia coli*: the HpaR repressor. *Nucleic Acids Res*; 31:6598-6609.
- [36] Gibson J, Harwood CS (2002). Metabolic diversity in aromatic compound utilization by anaerobic microbes. *Annu Rev Microbiol*; 56:345-369.
- [37] Gihring TM, Banfield JF (2001). Arsenite oxidation and arsenate respiration by a new *Thermus* isolate. *FEMS Microbiol Lett*; 204:335-340.
- [38] Gihring TM, Druschel GK, McCleskey RB, Hamers RJ, Banfield JF (2001). Rapid arsenite oxidation by *Thermus aquaticus* and *Thermus thermophilus*: field and laboratory investigations. *Environ Sci Technol*; 35:3857-3862.
- [39] Gireesh-Babu P, Chaudhari A (2012). Development of a broad-spectrum fluorescent heavy metal bacterial biosensor. *Mol Biol Rep*; 39(12):11225-11229.
- [40] Goering PL, Aposhian HV, Mass MJ, Cebrian M, Beck BD, Waalkes MP (1999). The enigma of arsenic carcinogenesis: Role of metabolism. *Toxicol Sci*; 49:5-14.
- [41] Griffiths E, Gupta RS (2007). Identification of signature proteins that are distinctive of the *Deinococcus-Thermus* phylum. *Int Microbiol*; 10:201-208.
- [42] Guo Z, Houghton JE (1999). PcaR-mediated activation and repression of *pca* genes from *Pseudomonas putida* are propagated by its binding to both the -35 and the -10 promoter elements. *Mol Microbiol*; 32:253-263.
- [43] Henne A, Brüggemann H, Raasch C, Wiezer A, Hartsch T, Liesegang H, Johann A, Lienard T, Gohl O, Martínez-Arias R, Jacobi C, Starkuviene V, Schlenczeck S, Dencker S, Huber R, Klenk HP, Kramer W, Merkl R, Gottschalk G, Fritz HJ (2004). The genome sequence of the extreme thermophile *Thermus thermophilus*. *Nat Biotechnol*; 22:547-553.
- [44] Hillson NJ, Hu P, Andersen GL, Shapiro L (2007). *Caulobacter crescentus* as a Whole-Cell Uranium Biosensor. *Appl and Environ Microbiology*; 73:7615-7621.
- [45] Huang WE, Wang H, Zheng H, Huang L, Singer AC, Thompson I, Whiteley AS (2005). Chromosomally located gene fusions constructed in *Acinetobacter* sp. ADP1 for the detection of salicylate. *Environ Microbiol*; 7:1339-1348.
- [46] Izzo V, Notomista E, Picardi A, Pennacchio F, Di Donato A (2005). The thermophilic archaeon *Sulfolobus solfataricus* is able to grow on phenol. *Res Microbiol*; 156:677-689.
- [47] Jackson CR, Harrison KG, Dugas SL (2005). Enumeration and characterization of culturable arsenate resistant bacteria in a large estuary. *Syst Appl Microbiol*; 28:727-734.
- [48] Ji G, Silver SJ (1992). Regulation and expression of the arsenic resistance operon from *Staphylococcus aureus* plasmid pI258. *Bacteriol*; 174:3684-3694.
- [49] Kieft TL, Fredrickson JK, Onstott TC, Gorby YA, Kostandarithes HM, Bailey TJ, Kennedy DW, Li SW, Plymale AE, Spadoni CM, Gray MS (1999). Dissimilatory reduction of Fe(III) and other electron acceptors by a *Thermus* isolate. *Appl Environ Microbiol*; 65:1214-1221.
- [50] Klomsiri C, Panmanee W, Dharmsthiti S, Vattanaviboon P, Mongkolsuk S (2005). Novel roles of *ohrR-ohr* in *Xanthomonas* sensing, metabolism, and physiological adaptive response to lipid hydroperoxide. *J Bacteriol*; 187:3277-3281.
- [51] Krell T, Lacala J, Guazzaronib ME, Buscha A, Jiménez HS, Filleta S, Reyes-Darías JA, Muñoz-Martínez F, Jiménez MR, Fontanaa CG, Duquea E, Seguraa A, Ramosa JL (2012).

- Responses of *Pseudomonas putida* to toxic aromatic carbon sources. Journal of Biotechnology; 160:25-32.
- [52] Kulkarni M, Chaudhari A (2007). Microbial remediation of nitro-aromatic compounds: an overview. J Environ Manage; 85(2):496-512.
 - [53] Kumarevel T, Tanaka T, Umehara T, Yokoyama S (2009). ST1710-DNA complex crystal structure reveals the DNA binding mechanism of the MarR family of regulators. Nucleic Acids Res; 37:4723-4735.
 - [54] Lee JW, Soonsanga S, Helmann JD (2007). A complex thiolate switch regulates the *Bacillus subtilis* organic peroxide sensor OhrR. Proc Natl Acad Sci USA; 104:8743-8748.
 - [55] Li Y, Hu Y, Zhang X, Xu H, Lescop E, Xia B, Jin C (2007). Conformational fluctuations coupled to the thiol-disulfide transfer between thioredoxin and arsenate reductase in *Bacillus subtilis*. Journal of biological chemistry; 282:11078-11083.
 - [56] Liao VH, Chien M, Tseng Y, Ou K (2006). Assessment of heavy metal bioavailability in contaminated sediments and soils using green fluorescent protein-based bacterial biosensors. Environmental Pollution; 146:17-23.
 - [57] Lin YF, Walmsley AR, Rosen BP (2006). An arsenic metallochaperone for an arsenic detoxification pump. PNAS; 103:15617-1562.
 - [58] Linde HJ, Notka F, Metz M, Kochanowski B, Heisig P, Lehn N (2000). *In vivo* increase in resistance to ciprofloxacin in *Escherichia coli* associated with deletion of the C-terminal part of MarR. Antimicrob Agents Chemother; 44:1865-1868.
 - [59] Lopez-Maury L, Sanchez-Riego AM, Reyes JC, Florencio FJ (2009). The glutathione /glutaredoxin system is essential for arsenate reduction in *Synechocystis* sp. strain PCC 6803. J Bacteriol; 191(11):3534-3543.
 - [60] Lovley DR (2003). Cleaning up with genomics: applying molecular biology to bioremediation. Nat Rev Microbiol; 1:35-44.
 - [61] Martin RG, Rosner JL (1995). Binding of purified multiple antibiotic-resistance repressor protein (MarR) to *mar* operator sequences. Proc Natl Acad Sci USA; 92:5456-5460.
 - [62] Mejäre M, Bulow L (2001). Metal-binding proteins and peptides in bioremediation and phytoremediation of heavy metals. Trends Biotechnol; 19:67-73.
 - [63] Meng YL, Liu Z, Rosen BP (2004). As(III) and Sb (III) uptake by GlpF and efflux by ArsB in *Escherichia coli*. J Biol Chem; 279(18):18334-18341.
 - [64] Messens J, Silver S (2006). Arsenate reduction: thiol cascade chemistry with convergent evolution. J Mol Biol; 362:1-17.
 - [65] Mitchell RJ, Gu MB (2004). Construction and characterization of novel dual stress-responsive bacterial biosensors. Biosensors and Bioelectronics; 19:977-985.
 - [66] Mukhopadhyay R, Rosen BP (2001). The phosphatase C(X)₅R motif is required for catalytic activity of the *Saccharomyces cerevisiae* Acr2p arsenate reductase. J Biol Chem; 276(37):34738-34742.
 - [67] Murota C, Matsumoto H, Fujiwara S, Hiruta Y, Miyashita S, Shimoya M, Kobayashi I, Hudock MO, Togasaki Norihiro RK, Sato N (2012). Arsenic tolerance in a *Chlamydomonas* photosynthetic mutant is due to reduced arsenic uptake even in light conditions. Planta; 236:1395-1403.
 - [68] Murphy JN, Saltikov CW (2009). The ArsR Repressor mediates arsenite-dependent regulation of arsenate respiration and detoxification operons of *Shewanella* sp. Strain ANA-3. Journal of Bacteriology; 191:6722-6731.
 - [69] Nayak M, Kotian A, Marathe S, Chakravorty D (2009). Detection of microorganism using biosensors-A smarter way towards detection techniques. Biosensors and Bioelectronics; 25:661-667.
 - [70] Newberry KJ, Fuangthong M, Panmanee W, Mongkolsuk S, Brennan RG (2007). Structural mechanism of organic hydroperoxide induction of the transcription regulator OhrR. Mol Cell; 28:652-664.
 - [71] Nimick DA, Moore JN, Dalby CE, Savka MW (1998). The fate of geothermal arsenic in the Madison and Missouri Rivers, Montana and Wyoming. Water Resources Research, v. 34:3051-3067.
 - [72] Ogawa N, McFall SM, Klem TJ, Miyashita K, Chakrabarty AM (1999). Transcriptional activation of the chlorocatechol degradative genes of *Ralstonia eutropha* NH9. J Bacteriol; 181:6697-6705.
 - [73] Oh SY, Shin JH, Roe JH (2007). Dual role of OhrR as a repressor and an activator in response to organic hydroperoxides in *Streptomyces coelicolor*. J Bacteriol; 189:6284-6292.

- [74] Okada U, Sakai N, Yao M, Watanabe N, Tanaka I (2006). Structural analysis of the transcriptional regulator homolog protein from *Pyrococcus horikoshii* OT3. *Proteins*; 63(4):1084-1086.
- [75] Ordóñez E, Van Belle K, Roos G, De Galan S, Letek M, Gil JA, Wyns L, Mateos LM, Messens J (2009). Arsenate reductase, mycothiol, and mycoredoxin concert thiol/disulfide exchange. *J Biol Chem*; 284(22):15107-15116.
- [76] Oremland RS, Stolz JF (2005). Arsenic, microbes and contaminated aquifers. *Trends Microbiol*; 13:45-49.
- [77] Ouhammouch M, Langham GE, Hausner W, Simpson AJ, El-Sayed NM, Geiduschek EP (2005). Promoter architecture and response to a positive regulator of archaeal transcription. *Mol Microbiol*; 56:625-637.
- [78] Pérez-Espino D, Tamames J, de Lorenzo V, Cánovas D (2009). Microbial responses to environmental arsenic. *Biometals*; 22:117-130.
- [79] Paitan Y, Biran I, Shechter N, Biran D, Rishpon J, Ron EZ (2004). Monitoring aromatic hydrocarbons by whole cell electrochemical biosensors. *Anal Biochem*; 335:175-183.
- [80] Patel PC, Goulhen F, Boothman C, Gault AG, Charnock JM, Kalia K, Lloyd JR (2007). Arsenate detoxification in a *Pseudomonad* hypertolerant to arsenic. *Arch Microbiol*; 187(3):171-183.
- [81] Peeters E, Charlier D (2010). The Lrp family of transcription regulators in archaea. *Archaea*; 30:750457. Doi: 10.1155/2010/750457.
- [82] Pennella MA, Shokes JE, Cosper NJ, Scott RA, Giedroc DP (2003). Structural elements of metal selectivity in metal sensor proteins. *Proc Natl Acad Sci USA*; 100:3713-3718.
- [83] Perera IC, Grove A (2010). Molecular mechanisms of ligand-mediated attenuation of DNA binding by MarR family transcriptional regulators. *J Mol Cell Biol*; 2(5):243-254.
- [84] Perera IC, Lee YH, Wilkinson SP, Grove A (2009). Mechanism for attenuation of DNA binding by MarR family transcriptional regulators by small molecule ligands. *J Mol Biol*; 390:1019-1029.
- [85] Qin J, Rosen BP, Zhang Y, Wang G, Franke S, Rensing C (2006). Arsenic detoxification and evolution of trimethylarsine gas by a microbial arsenite S-adenosylmethionine methyltransferase. *Proc Natl Acad Sci USA*; 103:2075-2080.
- [86] Ratnaike RN (2003). Acute and chronic arsenic toxicity. *Postgrad Med J*; 79:391-396.
- [87] Rieger PG, Meier HM, Gerle M, Vogt U, Groth T, Knackmuss HJ (2002). Xenobiotics in the environment: present and future strategies to obviate the problem of biological persistence. *J Biotechnol*; 94:101-123.
- [88] Rosen BP (2002). Biochemistry of arsenic detoxification. *FEBS Lett*; 529(1):86-92.
- [89] Rouanet C, Reverchon S, Rodionov DA, Nasser W (2004). Definition of a consensus DNA-binding site for PecS, a global regulator of virulence gene expression in *Erwinia chrysanthemi* and identification of new members of the PecS regulon. *J Biol Chem*; 279(29):30158-30167.
- [90] Saltikov CW, Olson BH (2002). Homology of *Escherichia coli* R773 *arsA*, *arsB*, and *arsC* genes in arsenic-resistant bacteria isolated from raw sewage and arsenic-enriched creek waters. *Appl Environ Microbiol*; 68:280-288.
- [91] Saridakis V, Shahinas D, Xu X, Christendat D (2008). Structural insight on the mechanism of regulation of the MarR family of proteins: high-resolution crystal structure of a transcriptional repressor from *Methanobacterium thermoautotrophicum*. *J Mol Biol*; 377:655-667.
- [92] Sato T, Kobayashi Y (1998). The *ars* operon in the skin element of *Bacillus subtilis* confers resistance to arsenate and arsenite. *J Bacteriol*; 180:1655-1661.
- [93] Say R, Yilmaz N, Denizli A (2003). Biosorption of cadmium, lead, mercury, and arsenic ions by the fungus *Penicillium purpurogenum*. *Sep Sci Technol*; 38:2039-2053.
- [94] Schell MA, Wender PE (1986). Identification of the *nahR* gene product and nucleotide sequences required for its activation of the *sal* operon. *J Bacteriol*; 166:9-14.
- [95] Shi W, Dong J, Scott RA, Ksenzenko MY, Rosen BP (1996). The role of arsenic-thiol interactions in metalloregulation of the *ars* operon. *J Biol Chem*; 271:9291-9297.
- [96] Singh VK, Xiong A, Usgaard TR, Chakrabarti S, Deora R, Misra TK, Jayaswal RK (1999). ZntR is an autoregulatory protein and negatively regulates the chromosomal zinc resistance operon *znt* of *Staphylococcus aureus*. *Mol Microbiol*; 33:200-207.
- [97] Sinha A, Kumar S, Khare SK (2013). Biochemical basis of mercury remediation and bioaccumulation by *Enterobacter* sp. EMB21. *Appl Biochem Biotechnol*; 169(1):256-267.
- [98] Stolz JF, Basu P, Santini JM, Oremland RS (2006). Arsenic and selenium in microbial metabolism. *Annu Rev Microbiol*; 60:107-130.
- [99] Thaw P, Sedelnikova SE, Muranova T, Wiese S, Ayora S, Alonso JC, Brinkman AB, Akerboom J, van der Oost J, Rafferty JB (2006). Structural insight into gene transcriptional regulation and effector binding by the Lrp/AsnC family. *Nucleic Acids Res*; 34:1439-1449.

- [100] Torres B, García JL, Díaz E (2004). Plasmids as tools for containment. In: Phillips G, Funnell B (eds) *Plasmid Biology*. ASM Press, Washington DC, pp 589-601.
- [101] Tropel D, van der Meer JF (2004). Bacterial Transcriptional regulators for degradation pathways of aromatic compounds. *Microbiology and Molecular Biology Reviews*; 68(3):474-500.
- [102] Turner JS, Glands PD, Samson ACR, Robinson NJ (1996). Zn²⁺-sensing by the cyanobacterial metallothionein repressor SmtB: different motifs mediate metal-induced protein-DNA dissociation. *Nucleic Acids Res*; 19:3714-3721.
- [103] Valls M, de Lorenzo V (2002). Exploiting the genetic and biochemical capacities of bacteria for the remediation of heavy metal pollution. *FEMS Microbiol Rev*; 26:327-338.
- [104] VanZile ML, Chen X, Giedroc DP (2002). Allosteric negative regulation of *smt* O/P binding of the zinc sensor, SmtB, by metal ions: a coupled equilibrium analysis. *Biochemistry*; 41:9776-9786.
- [105] Volant A, Desoeuvre A, Casiot C, Lauga B, Delpoux S, Morin G, Personné JC, Héry M, Elbaz-Poulichet F, Bertin PN, Bruneel O (2012). Archaeal diversity: temporal variation in the arsenic-rich creek sediments of Carnoulès Mine, France. *Extremophiles*; 16:645-657.
- [106] Wilkinson SP, Grove A (2005). Negative cooperativity of uric acid binding to the transcriptional regulator HucR from *Deinococcus radiodurans*. *J Mol Biol*; 350:617-630.
- [107] Wong MD, Lin YF, Rosen BP (2002). The soft metal ion binding sites in the *Staphylococcus aureus* pl258 CadC Cd(II)/Pb(II)/Zn(II)-responsive repressor are formed between subunits of the homodimer. *J Biol Chem*; 277:40930-40936.
- [108] Wu J, Tisa LS, Rosen BP (1992). Membrane topology of the ArsB protein, the membrane subunit of an anion-translocating ATPase. *J Biol Chem*; 267:12570-12576.
- [109] Xiong A, Gottman A, Park C, Baetens M, Pandza S, Matin A (2000). The EmrR protein represses the *Escherichia coli* *emrRAB* multidrug resistance operon by directly binding to its promoter region. *Antimicrob Agents Chemother*; 44:2905-2907.
- [110] Zegers I, Martins JC, Willem R, Wyns L, Messens J (2001). Arsenate reductase from *S. aureus* plasmid pl258 is a phosphatase drafted for redox duty. *Nature Structural Biology*; 8:843-847.
- [111] Zhou T, Radaev S, Rosen BP, Gatti DL (2000). Structure of the ArsA ATPase: the catalytic subunit of a heavy metal resistance pump. *EMBO J*; 19:1-8.

Chapter 2

IDENTIFICATION AND PHYSICOCHEMICAL, STRUCTURAL AND FUNCTIONAL CHARACTERIZATION OF BldR2 FROM *SULFOLOBUS SOLFATARICUS* P2

- Identification and physicochemical characterization of BldR2 from *Sulfolobus solfataricus*, a novel archaeal member of the MarR transcription factor family

Reprinted with permission from the Biochemistry

- Characterization of ligand and DNA binding abilities of BldR2 and BldR

Identification and Physicochemical Characterization of BldR2 from *Sulfolobus solfataricus*, a Novel Archaeal Member of the MarR Transcription Factor Family

Gabriella Fiorentino,^{*,†} Immacolata Del Giudice,[†] Simonetta Bartolucci,[†] Lorenzo Durante,[‡] Luigi Martino,[§] and Pompea Del Vecchio^{*,‡}

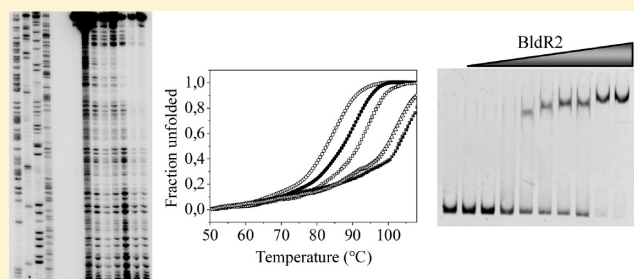
[†]Department of Structural and Functional Biology, University of Naples Federico II, Edificio 7, via Cinthia, 80126 Naples, Italy

[‡]Department of Chemistry "Paolo Corradini", University of Naples Federico II, via Cinthia, 80126 Naples, Italy

[§]Randall Division of Cell and Molecular Biophysics, King's College London, New Hunt's House, Guy's Campus, London SE1 1UL, U.K.

S Supporting Information

ABSTRACT: The multiple antibiotic resistance regulators (MarR) constitute a family of ligand-responsive transcriptional regulators abundantly distributed throughout the bacterial and archaeal domains. Here we describe the identification and characterization of BldR2, as a new member of this family, in the archaeon *Sulfolobus solfataricus* and report physiological, biochemical, and biophysical investigation of its stability and DNA binding ability. Transcriptional analysis revealed the upregulation of *BldR2* expression by aromatic compounds in the late-logarithmic growth phase and allowed the identification of *cis*-acting sequences. Our results suggest that BldR2 possesses in solution a dimeric structure and a high stability against both temperature and chemical denaturing agents; the protein binds site specifically to its own promoter, *Sso1082*, with a micromolar binding affinity at two palindromic sites overlapping TATA-BRE and the transcription start site. Benzaldehyde and salicylate, ligands of MarR members, are antagonists of binding of DNA by BldR2. Moreover, two single-point mutants of BldR2, R19A and A65S, properly designed for obtaining information about the dimerization and the DNA binding sites of the protein, have been produced and characterized. The results point out an involvement of BldR2 in the regulation of the stress response to aromatic compounds and point to arginine 19 as a key amino acid involved in protein dimerization, while the introduction of serine 65 increases the DNA affinity of the protein, making it comparable with those of other members of the MarR family.



Archaea, the third domain of life, are microorganisms adapted to grow in extreme environments with regard to temperature, pH, ionic strength, and high concentrations of detergents and organic solvents.¹ They are an evolutionary mosaic, being more similar to eukaryotes with respect to the macromolecular machinery and more similar to bacteria with respect to metabolic systems and genome organization.^{2–4} For example, with regard to the transcription machinery, in most cases, homologues of bacterial regulators function in the context of the archaeal basal transcriptional apparatus, which resembles that of Eukarya. Archaea possess a TATA box promoter sequence, a TATA box-binding protein (TBP), a homologue of the transcription factor TFIIB (TFB), and an RNA polymerase (RNAP) containing between 8 and 13 subunits.^{5,6}

Archaea have been of interest to many protein chemists over the years, as models for understanding the molecular basis of adaptation to extreme conditions. In the case of adaptations to extremes of pH, salinity, and pressure, it has been proven that membrane components and protective small molecules may play important roles.⁷ However, with regard to temperature

adaptations, the cellular components themselves, namely the proteins, have to achieve thermostability.^{8,9} Like all other living cells, archaea are also able to defend against subtle changes to environmental conditions; they own in their genomes finely regulated biochemical pathways for detoxification as well as different regulative sequences responsive to stress agents.^{10–13} The multiple antibiotic resistance regulators (MarR) constitute a family of ligand-responsive transcriptional regulators that are distributed throughout the bacterial and archaeal domains and include proteins critical for the control of virulence factor production, the response to oxidative stress, and the regulation of the catabolism of environmental aromatic compounds. They are also involved in the regulation of mechanisms of resistance to multiple antibiotics, organic solvents, household disinfectants, and pathogenic factors. MarR homologues are dimeric proteins that have a low level of sequence identity and a triangular shape; they

Received: February 4, 2011

Revised: June 28, 2011

Published: June 30, 2011

bind to their cognate palindromic or pseudopalindromic DNA as homodimers, resulting in either transcriptional repression or activation.¹⁴ The DNA binding domain of MarR proteins is a conserved winged helix–turn–helix motif¹⁵ with the two wings located at the corners of the triangle. Another common feature of MarR members is their ability to interact with specific ligands and, upon binding, to modulate DNA recognition.¹⁶ Crystal structures of several MarR regulators have been obtained, either as apoproteins, in complex with the cognate DNA, or with various effectors, greatly contributing to the elucidation of the mechanistic basis of DNA and/or ligand binding. However, the identification of key residues involved in binding as well as those contributing to protein stability and/or dimerization has been reported only in a few cases.^{17–19}

In the archaeal domain, the crystal structures of four transcription factors, ST1710 (or StEmrR) from *Sulfolobus tokodaii*,^{20,21} MTH313 from *Methanobacterium thermoautotrophicum*,²² PH1061 from *Pyrococcus horikoshii* OT3,²³ and BldR from *Sulfolobus solfataricus*,²⁴ have been determined. The overall structure of all of these proteins is typical of the MarR family, particularly for elements taking part in the DNA-binding domain, while the most important differences are found in the dimerization domain. The structures of MTH313 and ST1710 complexed with salicylate as the ligand revealed conservation of the ligand binding pocket.^{17,22} Furthermore, it has been proposed that the ability to act as activators or repressors could not be related to a particular DNA binding mechanism, but rather to the position of the binding site on the target DNA.²⁴ Comparison of the thermophilic StEmrR with mesophilic MarR showed an increase in the number of salt bridges on the protein surface predicted to be important in increasing the thermostability.²¹

Some of us identified BldR as part of an operon-like structure conserved in most archaea. Functional characterization demonstrated that BldR is a transcriptional activator involved in the detoxification of aromatic compounds.^{25,26} The *S. solfataricus* genome contains an additional ORF, *Sso1082*, that encodes another putative MarR protein of 154 amino acids that is 35% identical in sequence to BldR of the same organism.

In this work, we have performed a physiological and biochemical analysis of BldR2 with the aim of enhancing our knowledge of the role of MarR proteins in the archaeal domain. Our results showed that the gene is transcriptionally regulated. Furthermore, BldR2 is a homodimer that binds specifically to its own promoter in a region that overlaps with the sequences recognized by the basal transcription machinery. In the presence of benzaldehyde and salicylate, ligands of MarR members, BldR2 dissociates from its promoter. In this study, we also report interesting results in terms of conformational stability and DNA binding properties of BldR2 through biophysical and biochemical measurements. Moreover, to provide further insights into thermal stability and DNA binding molecular mechanisms of BldR2, we used guided mutagenesis based on the structure of the close relatives BldR²⁴ and ST1710²¹ and identified two residues, Arg19, possibly involved in dimer stabilization, and Ala65, located in the DNA binding domain. BldR has in the same position a serine residue directly involved in DNA binding.²⁴ These two residues, Arg19 and Ala65, were substituted in two different mutants with Ala and Ser, respectively, and a complete characterization was conducted in parallel with wild-type BldR2.

MATERIALS AND METHODS

***S. solfataricus* Cultivation and Preparation of Genomic DNA and Total RNA.** *S. solfataricus* P2 was grown aerobically at 82 °C in 100 mL of medium described by Brock supplemented with 0.1% (w/v) yeast extract and 0.1% (w/v) casamino acids²⁷ and buffered at pH 3.5. In some cases, benzaldehyde, sodium salicylate, and benzyl alcohol were added to final concentrations of 1, 0.35, and 4 mM, respectively, after dilution of an exponentially growing culture up to an A_{600} of 0.08 optical density (OD) unit. Cells were grown to ~0.3 and ~0.7 OD₆₀₀ unit, corresponding to midlogarithmic and late-logarithmic phases, respectively, and harvested by centrifugation at 4000g for 10 min. Genomic DNA and total RNAs were prepared following reported procedures.²⁸

In Vivo Response to Aromatic Compounds. RNAs (10 µg) extracted from cells grown under different conditions were electrophoretically separated in a 1.5% agarose gel containing 10% formaldehyde and transferred to nylon filters (GE Healthcare). Hybridization was conducted as described by Cannio et al.,²⁸ using the *Sso1082* and *Sso1352* genes²⁵ and rRNA 16S as probes. The experiments were performed in duplicate. Signals were visualized by autoradiography and quantified with a densitometric analysis using a Personal Fx phosphorimager and Quantity One (Bio-Rad).

Primer Extension Analysis of the Transcription Start Site. To determine the first transcribed nucleotide, total RNA extracted from cells grown in the presence or absence of benzaldehyde and harvested at 0.3 OD₆₀₀ unit was subjected to primer extension analysis as described by Limauro et al.,²⁹ using primer *Sso1082+100Rv* (5'-GGC CTA TTT GCT CAA GAG CC-3'), annealing from position 100 bp in the *Sso1082* gene. The same primer was used to produce a sequence ladder by using the f-Mol DNA cycle sequencing system (Promega), according to the manufacturer's instructions, to locate the products on a 6% urea–polyacrylamide gel.

Heterologous Expression of *Sso1082* and Purification of the Recombinant Protein. The gene encoding *Sso1082* from *S. solfataricus* P2 was amplified by polymerase chain reaction (PCR) from genomic DNA, using *Pfx* DNA polymerase. Two different upstream primers both containing the *NdeI* restriction site (underlined) (*Sso1082up*, 5'-GGA TTT TGT GAG TTCATATGA TG-3'; *Sso1082Fw*, 5'-GTT AGA TAT CTA CAT ATG ATA TTA GC-3') were designed on the basis of two different putative translation start sites. In particular, *Sso1082Ssoup* anneals to the ATG annotated on the genome³⁰ while *Sso1082Fw* anneals to another putative start codon located 45 nucleotides downstream and deduced from both a transcriptional analysis³¹ and our primer extension. The common downstream primer *Sso1082 Rev* (5'-GCT TTA AGA CTC GAG TAG TTA GG-3') introduces a stop codon and the *XhoI* restriction site (underlined in the sequence). Amplified fragments were purified, digested with appropriate restriction enzymes, and cloned in the pET30a *NdeI*- and *XhoI*-modified vector, generating pET30*Sso1082long* and pET30*Sso1082*, respectively. The sequences of the two cloned fragments were shown to be identical to those available on the *S. solfataricus* P2 genome (<http://www-archbac.u-psud.fr/projects/sulfolobus/>).

Escherichia coli BL21-CodonPlus (DE3) cells (Stratagene) transformed with pET30*Sso1082long* and pET30*Sso1082* were used for recombinant protein expression.

Cells transformed with pET30Sso1082 were grown in 1 L of Luria-Bertani medium containing kanamycin (100 $\mu\text{g}/\text{mL}$) and chloramphenicol (33 $\mu\text{g}/\text{mL}$). When the culture reached an A_{600} of 0.5 OD unit, protein expression was induced by the addition of 0.5 mM IPTG and the bacterial culture grown for an additional 6 h. Cells were harvested by centrifugation, and pellets were lysed by sonication in 30 mL of lysis buffer [50 mM Tris-HCl (pH 8.0) and 1 mM phenylmethanesulfonyl fluoride] in an ultrasonic liquid processor (Heat system Ultrasonic Inc.). The lysate was centrifuged at 30000g for 60 min (SW41 rotor, Beckman). The supernatant was heated to 80 °C for 10 min, and denatured proteins were precipitated by centrifugation at 20000g for 20 min at 4 °C.

The supernatant was loaded onto a heparin column (5 mL, HiTrap heparin, GE Healthcare) equilibrated in 50 mM Tris-HCl (pH 8.0) (buffer A) connected to an AKTA Explorer system (GE Healthcare). After a washing step with buffer A, elution was conducted with 40 mL of a KCl gradient (0 to 0.8 M). Fractions containing the BldR2 protein were pooled, concentrated, dialyzed, and loaded onto a Superdex 75 column (26 cm \times 60 cm, GE Healthcare) equilibrated in 50 mM Tris-HCl (pH 8.0) and 0.2 M KCl (buffer B) at a flow rate of 2 mL/min. Fractions were collected and analyzed by SDS-PAGE to detect the BldR2 protein. These fractions were pooled, concentrated by ultrafiltration using a YM10 membrane (Millipore), dialyzed against buffer A, and stored at 4 °C.

Construction, Expression, and Purification of R19A and A65S Mutants. Single-point mutations in the *BldR2* gene were produced with the Quick Change Site-Directed Mutagenesis Kit (Stratagene) that utilizes *PfuUltra* high-fidelity DNA polymerase and primers complementary to the coding and noncoding template sequences.

To generate the R19A and A65S mutants, the forward primers (*FwSso1082R19A*, 5'-TTGGGATCTAATAACTGGACTGACAGCAAAGATAAATAAGGATACAGATAAG-3'; *FwSso1082A65S*, 5'-GCTGAAAAATATATGCTGACAAAGTCGGGATTAAGTAGCATCATT-3') with their complementary reverse primers were used (underlined letters indicate the base pair mismatch) and reactions performed according to the manufacturer's instructions. The mutagenesis products were transformed into XL-1 Blue Cells. Single colonies were selected on LB plates containing kanamycin, and isolated plasmidic DNAs were sequenced at Eurofins MWG Operon.

Plasmids pET30R19A and pET30A65S containing the desired mutations were used to transform BL21 Codon plus (DE3) competent cells. The best growth conditions for gene expression were determined both for cells transformed with pET30R19A and for cells transformed with pET30A65S: growth to an OD_{600} of ~ 0.5 and ~ 0.8 in LB medium supplemented with kanamycin (50 $\mu\text{g}/\text{mL}$) and chloramphenicol (33 $\mu\text{g}/\text{mL}$) at 37 °C, respectively, followed by induction for 6 h with 0.5 mM IPTG.

The purification of the mutant proteins was conducted in a manner similar to that already described for the wild-type BldR2 protein.

Computational Methods. To establish similarities among the sequences of proteins of interest and the sequences of the SwissProt Data Bank, computational analysis was performed at <http://www-arch-bac.u-psud.fr/projects/sulfolobus/> or <http://cmr.jcvi.org/cgi-bin/CMR/GenomePage.cgi?org=ntss02>, providing access to the genome of *S. solfataricus* P2.

The multiple-sequence alignment was obtained using T-COFFEE.³² The three-dimensional (3D) model of BldR2 has

been built with EsysPred3D using the BldR 3D structure as a template.³³

Analytical Methods for Protein Characterization and Sample Preparation. Protein concentrations were determined by the method of Bradford, using BSA as the standard.³⁴ Protein homogeneity was estimated by SDS-PAGE [12.5% (w/v) gels]. To determine the native molecular mass of the proteins, the purified proteins at different concentrations (0.25, 0.5, 1.5, and 3.0 mg/mL) were applied in a volume of 100 μL to an analytical Superdex PC75 column (3.2 cm \times 30 cm) connected to an AKTA Explorer system (GE Healthcare) alternatively equilibrated with buffer B, with 20 mM sodium phosphate (pH 7.5) and 0.2 M KCl, or with the same buffer used for EMSA (see below), at a flow rate of 0.04 mL/min. The column was calibrated in the different buffers using a set of gel filtration markers (low range, GE Healthcare), including bovine serum albumin (67.0 kDa), ovalbumin (43.0 kDa), chymotrypsinogen A (25.0 kDa), and RNase A (13.7 kDa). The molecular mass of the protein and mutants was also determined using electrospray mass spectra recorded on a Bio-Q triple quadrupole instrument (Micromass, Thermo Finnigan, San Jose, CA).

Protein solutions for spectroscopic analyses were prepared in a 20 mM sodium phosphate buffer (pH 7.5), and the concentration was determined by UV spectra using a theoretical, sequence-based extinction coefficient of 22920 $\text{M}^{-1} \text{cm}^{-1}$ calculated at 280 nm for the dimeric protein.³⁵ A commercial 8 M GuHCl solution from Sigma was used as a denaturant solvent. Protein solutions for CD and fluorescence measurements were exhaustively dialyzed by using Spectra Por membranes (molecular weights of 15000–17000) against buffer solution at 4 °C. The water used for buffer and sample solutions was doubly distilled. The pH was measured at 25 °C with a Radiometer model PHM93 pH meter.

Samples for GuHCl-induced denaturation experiments were prepared with increasing amounts of denaturing agent. Each sample was mixed by vortexing and incubated at 4 °C for 1 day. Longer incubation times produced identical spectroscopic signals.

Cloning of *Sso1082* Promoter Regions. Two different regions upstream of the ORF *Sso1082* were amplified by PCR amplification on *S. solfataricus* P2 genomic DNA: the first of 230 bp was obtained using the primer pair *Sso1082fw-130* (5'-ATT AGG ATA TAG ATC TCG TTT ACG A-3') and *Sso1082+100Rv*. The *Sso1082prFw* primer anneals starting at position -130 with respect to the transcription initiation site as determined by primer extension analysis.

A second smaller region of 164 bp was amplified with the pair *Sso1082fw-130* (see above) and *Sso1082+34Rv* (5'-CCCAAAC-TTCTGAGTACTTTGTAG-3') annealing from position +34 in the *Sso1082* gene.

The fragments were cloned in the pGEM T-easy vector (Promega) and TopoTA (Invitrogen) to give *Sso1082L* (large) and *Sso1082S* (small), respectively. The insertion and correct sequence of the PCR products were verified by DNA sequencing.

DNA Binding. Binding of BldR2, R19A, and A65S to the putative regulatory region *Sso1082L* was measured by an electrophoretic mobility shift assay (EMSA) using a biotin-labeled PCR fragment amplified from the pGEM T-easy vector using the *Sso1082Fw-130* and the *Sso1082+100Rv* primers (see above). The amplified DNA fragment was labeled at the 3'-OH end with the Biotin 3' End DNA labeling kit (Thermo Scientific), according to the manufacturer's protocol.

EMSA reaction mixtures were set up in a final volume of 15 μ L and contained 5 nM DNA, 1 μ g of poly(dI-dC), and varying amounts of proteins in binding buffer [25 mM Tris-HCl (pH 8), 50 mM KCl, 10 mM MgCl₂, 1 mM dithiothreitol, and 5% glycerol]. The mixtures were incubated at 60 °C for 20 min and run onto a nondenaturing 5% polyacrylamide gel (Bio-Rad) in 1 \times TBE at 80 V. The probes were transferred onto a positively charged nylon membrane (Hybond-XL, GE Healthcare, Uppsala, Sweden) with a blotting apparatus (Bio-Rad, Hercules, CA) and then detected with the Chemiluminescent Nucleic Acid Detection Module Kit (Thermo Scientific) according to the kit protocol.

To determine the dissociation constants of BldR2, R19A, and A65S with respect to the *Sso1082* promoter, the *Sso1082S* regulatory region (150 nM) was incubated with increasing amounts of the purified proteins and the complexes were separated as described above; after electrophoresis, gels were directly stained with SYBR green (nucleic acid gel stain, Invitrogen). At least two independent experiments were performed in duplicate. In particular, protein concentrations (in dimer units) ranged from 1.0 to 90 μ M for BldR2 and R19A and from 0.05 to 1.5 μ M for A65S.

Densitometric data were obtained with Quantity One (Bio-Rad) and manipulated to calculate the fractional complex formation (that is the ratio between the density of the retarded band and the total density, reported in percent). These values were analyzed by fitting the binding isotherm to the Hill equation in GraphPad Prism 5.0.

To analyze the effect of benzaldehyde and salicylate on the properties of BldR2, R19A, and A65S binding to the *Sso1082S* promoter, EMSAs were performed via preincubation of the proteins, at concentrations similar to their apparent K_d values, with 3, 5, 10, 20, 30, 50, and 100 mM sodium salicylate or benzaldehyde. Gels were processed and visualized as described above.

DNase I Footprinting. A probe containing the promoter region of the *Sso1082* gene was produced by PCR using a combination of the *Sso1082* *ftR* primer (5'-CGAATTCGCCCTTGGGTTTGAAG-3') designed on the basis of the *Sso1082* promoter (bold) and pTopo sequences, respectively, starting at +34 bp from the transcription initiation site, and *Sso1082*-184 (5'-CCATATTTATAATCTCTACA-3') as a second primer; the latter was labeled at the 5' end with T4 polynucleotide kinase and [γ -³²P]ATP. The labeled PCR product of 231 bp (~40 nM) was incubated with 5–10 μ g of pure BldR2, R19A, and A65S at 60 °C in binding buffer (see above) and digested with 1 unit of DNase I (Ambion) for 1 min at 37 °C. Subsequent steps were performed as described by Fiorentino et al.²⁵ Labeled primer was as also used to generate a dideoxynucleotide sequence ladder with the Promega f-Mol DNA sequencing system using *Sso1082S*-Topo (see above) as the template and following the manufacturer's instructions.

Circular Dichroism and Fluorescence Measurements. CD spectra were recorded with a Jasco J-715 spectropolarimeter equipped with a Peltier-type temperature control system (model PTC-348WI). The molar ellipticity per mean residue, $[\theta]$ in degrees square centimeters per decimole, was calculated from the equation $[\theta] = ([\theta]_{\text{obs}} \text{mrw}) / (10lC)$, where $[\theta]_{\text{obs}}$ is the ellipticity measured in degrees, mrw is the mean residue molecular weight (117.6) for protein BldR2, C is the protein concentration in grams per milliliter, and l is the optical path length of the cell in centimeters. Cells with path lengths of 0.1 and 1 cm were used in the far-UV and near-UV regions, respectively. CD spectra were recorded with a time constant of 4 s, a 2 nm bandwidth, and a

scan rate of 20 nm/min; the signal was averaged over at least three scans and baseline corrected by subtraction of a buffer spectrum. Spectra were analyzed for secondary structure amount according to the CDSSTR method³⁶ using Dichroweb.³⁷ The GuHCl-induced denaturation curves, at a fixed constant temperature of 25 °C, were obtained by recording the CD signal at 230 nm for the samples containing increasing amounts of GuHCl. Finally, the thermal unfolding curves were recorded in the temperature mode, by following the change in the CD signal at 222 nm with a scan rate of 1.0 °C/min. Fluorescence measurements were performed with a JASCO FP-750 apparatus equipped with a circulating water bath to keep the cell holders at a constant temperature of 25 °C. The denaturant-induced unfolding curves were obtained by recording changes in both fluorescence intensity and fluorescence maximal wavelength as a function of GuHCl concentration. The excitation wavelength was set to 290 nm, and the experiments were performed by using a 1 cm sealed cell and a 5 nm emission slit width and corrected for the background signal. The change in fluorescence intensity at 336 nm was recorded to monitor the unfolding transition. The protein concentration was kept constant at 2.4 μ M.

Analysis of the Denaturant-Induced Unfolding Transitions. For comparison of the denaturant-induced unfolding curves obtained in both the CD and fluorescence experiments, the curves were normalized reporting the fraction of unfolded protein (f_U) as a function of the concentration of the denaturing agent.

Thermodynamic parameters for the denaturant-induced unfolding were determined by analyzing the transition curves on the basis of a simple two-state model for dimeric proteins. In the equilibrium, only folded dimeric protein N_2 and unfolded monomers U exist. At any point in the denaturation reaction, the equilibrium constant K_U was calculated according to the model

$$K_U = [U]^2/[N_2] = 2P_t(f_U^2)/(1 - f_U) \quad (1)$$

in which P_t is the total molar concentration of protein monomers. The midpoint of the unfolding transition, C_m , was calculated using the equation

$$C_m = [RT \ln(P_t) + \Delta G_U(H_2O)]/m$$

The unfolding Gibbs energies were calculated using the relation

$$\Delta G_U = -RT \ln K_U \quad (2)$$

where R is the gas constant and T the absolute temperature. The linear dependence of the Gibbs energy of unfolding on the denaturant concentration is given by

$$\Delta G_U = \Delta G_U(H_2O) + m[D] \quad (3)$$

where $\Delta G_U(H_2O)$ is the extrapolated Gibbs energy of unfolding in the absence of denaturant³⁸ and m is the cooperativity parameter.³⁹ Values of $\Delta G_U(H_2O)$ were obtained directly by fitting the unfolding curves to eqs 1–3.

RESULTS

Transcriptional Analysis of *Sso1082*. Northern blot experiments were performed to verify the transcription of the *Sso1082* gene in two growth phases and to analyze transcription in cells grown in the presence of aromatic compounds, benzaldehyde, benzyl alcohol, and salicylate; these compounds were already demonstrated to inhibit cell growth at 1.5, 4, and 0.5 mM,

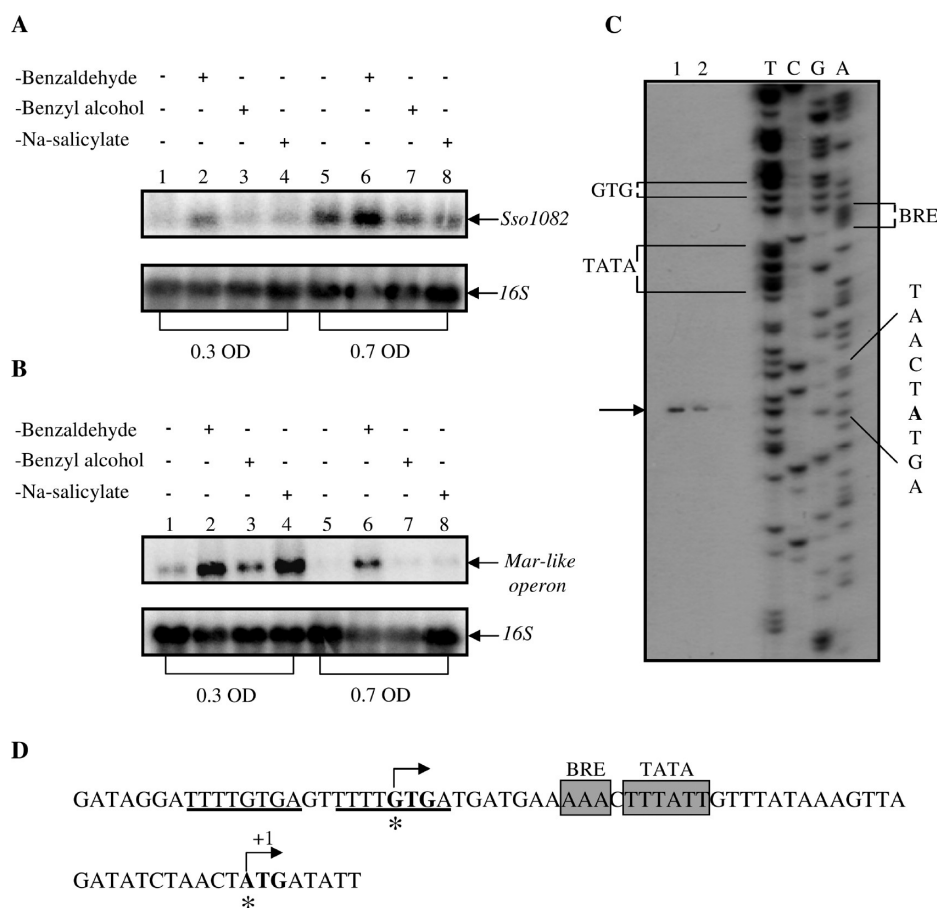


Figure 1. Transcriptional analysis of the *Sso1082* gene. Northern blot analysis of *Sso1082* (A) and *Mar-like operon* (B) mRNAs. Total RNA was prepared from *S. solfataricus* cells grown in the presence of different aromatic compounds and harvested in exponential (lanes 1–4) and stationary (lanes 5–8) growth phases: lanes 1 and 5, untreated control cells; lanes 2 and 6, cells grown in the presence of 1 mM benzaldehyde; lanes 3 and 7, cells grown in the presence of 4 mM benzyl alcohol; lanes 4 and 8, cells grown in the presence of 0.35 mM salicylate. The filters were probed with the *Sso1082* (A) and *Sso1352* (B) genes. Amounts of the mRNAs were normalized to 16S rRNA. The experiments were performed in duplicate. (C) Primer extension analysis of the *Sso1082* promoter region. Total RNA was prepared from cells grown in the presence (lane 1) or absence (lane 2) of benzaldehyde and harvested in the exponential growth phase. Primer-extended products were separated by electrophoresis under denaturing conditions alongside sequencing reactions with the same primer. (D) Promoter sequence analysis. The mapped transcription/translation start site (+1) is highlighted in bold; TBP and TFB binding sites are boxed. The initiation codon as annotated on the *S. solfataricus* genome is in bold and TSS as determined by Wurtzel are indicated by an asterisk.

respectively, and to affect *MarR-like operon* transcription.²⁵ The *BldR2* expression pattern was compared to that of the *Mar-like operon* and the 16S rRNA. *BldR2* mRNA revealed a single hybridization band under all the conditions tested with a molecular transcript of ~400 bp, which is slightly lower than that deduced from the gene sequence (462 bp) but is in accordance with a monocistronic transcript (Figure 1A). A densitometric analysis (Figure S1 of the Supporting Information) revealed that the level of *Sso1082* mRNA was ~3-fold higher in cells grown in the late exponential growth phase in comparison with that of cells grown in the early exponential growth phase (Figure 1A and Figure S1 of the Supporting Information, lanes 1 and 5).

Furthermore, the level of the transcript increased ~2-fold in cells grown in the presence of 1 mM benzaldehyde in comparison with nontreated cells in both growth phases (Figure 1A and Figure S1 of the Supporting Information, lanes 1, 2, 5, and 6). A weak induction could be observed when challenging cells with 4 mM benzyl alcohol (Figure 1A and Figure S1 of the Supporting Information, lanes 1–3 and 5–7) or 0.35 mM salicylate (Figure 1A and Figure S1 of the Supporting Information, lanes 1–4 and 5–8). The amounts of total cellular RNA were

comparable in all the experiments, which could be judged by hybridization of the same filters with the 16S rRNA gene. Taken together, these results are evidence that the level of *Sso1082* expression increases in a later growth stage with respect to that of the *Mar-like operon* and responds to stress by aromatic drugs (Figure 1B).

To determine the transcription start site of *Sso1082*, a primer extension analysis was undertaken on RNAs prepared from cells grown in the presence or absence of benzaldehyde. Unexpectedly, the results presented in Figure 1C reveal that the transcription start site corresponds to an adenosine located 45 bp downstream of the GTG start codon annotated on the *S. solfataricus* P2 genome.³⁰ Interestingly, a recent report on the *S. solfataricus* transcriptome has revealed that *Sso1082* transcription starts at two major positions: one is located in position –1 relative to the beginning of the ORF annotated on the genome and a second that coincides with the one we mapped and overlaps a putative ATG start codon; the two transcripts would produce proteins in the same frame differing for 15 amino acids at the N-terminus.³¹ It was also found that the extent of transcription from the downstream start site was >10-fold higher.³¹

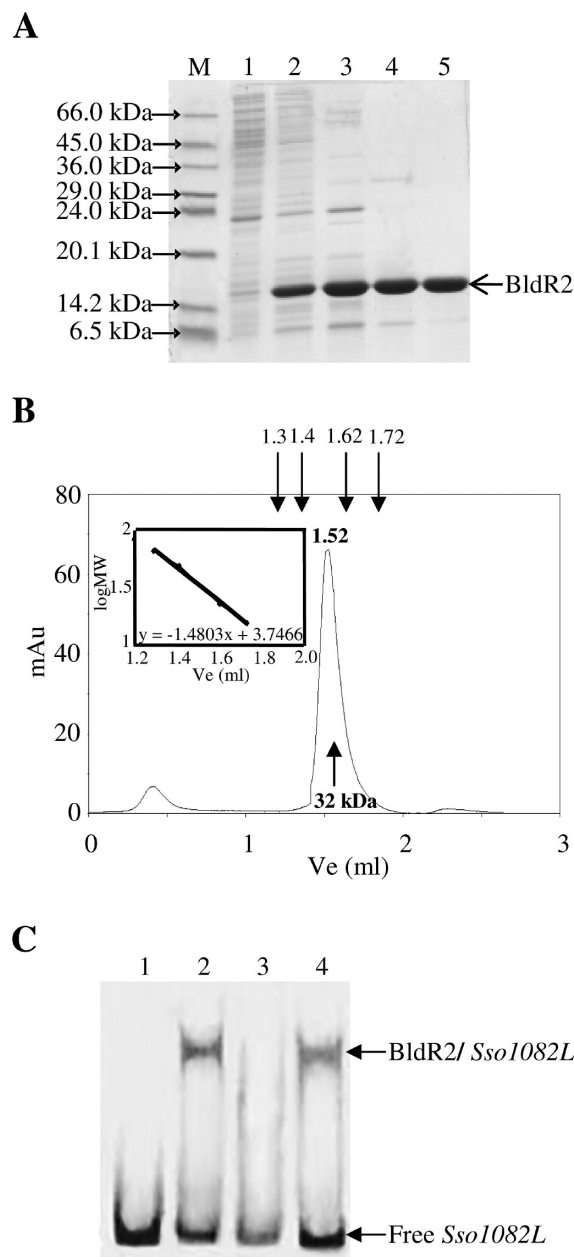


Figure 2. Analysis of recombinant BldR2. (A) SDS–PAGE of the purification steps of recombinant BldR2: lane M, molecular mass markers; lane 1, crude extract; lane 2, heat-treated cell extract; lane 3, fraction from heparin chromatography; lane 4, fraction from exclusion molecular chromatography; lane 5, purified BldR2. (B) Elution profile of purified BldR2 from gel filtration on a Superdex PC75 column. Recombinant BldR2 is eluted at 1.52 mL corresponding to a molecular mass of 32 kDa. Arrows indicate the elution volumes of the protein standards in the relative calibration of the column. (C) EMSA. *Sso1082L* (5 nM) and BldR2 incubated in the presence of specific and nonspecific competitors: lane 1, labeled DNA fragment; lane 2, DNA probe and BldR2 (15 μ M); lane 3, *Sso1082L*, BldR2, and 1250 nM nonlabeled *Sso1082L* fragment; lane 4, *Sso1082L*, BldR2, and 1250 nM nonlabeled *gfp* fragment.

However, under all of our growth conditions, we detected only the shorter transcript. It is tempting to speculate that under our experimental conditions transcription and translation start sites could coincide, giving rise to a single transcript, and in vivo, a

mature protein could be also translated starting from a Met residue located downstream from that found in the genome annotation.

The hypothesis is strengthened by the fact that the transcriptome study proved that leaderless translation is the preferred strategy in *S. solfataricus*.³¹ On the basis of the position of the downstream transcription start site, it was possible to identify TATA box and BRE sequences perfectly matching with the consensus, located downstream of the predicted GTG initiation codon (Figure 1D). The results depicted in Figure 1C confirm the relative increase in the level of mRNA upon benzaldehyde induction.

Heterologous Expression and Characterization of BldR2.

Both putative *Sso1082* genes predicted from genome annotation²⁶ and from the transcriptional analysis³¹ were cloned in pET30 and expressed in *E. coli*; interestingly, BldR2 was expressed in *E. coli* as a soluble protein only when the corresponding gene was cloned from the downstream start codon. All of our attempts to produce a soluble protein from the gene sequence as annotated in the *S. solfataricus* genome failed;³⁰ the protein purified from inclusion bodies was also found to be very sensitive to protease degradation at its N-terminus. Edman sequencing revealed, in fact, the absence of the first 13 amino acids (data not shown).

Hence, in the following description, we refer to BldR2 as the smaller protein translated from the downstream transcription and translation start site, and the reported characterization has been performed on this protein. BldR2 was purified to homogeneity (Figure 2A), taking advantage of its intrinsic properties, including its thermostability, its putative DNA binding capability, and its small size. Active fractions were selected on the basis of their ability to shift their own promoter region, *Sso1082L*, spanning positions –130 to 100 relative to its transcription and translation initiation codon in an EMSA (see also below). From 1 L of *E. coli* culture, it was possible to obtain up to 20 mg of pure BldR2. The molecular mass of the recombinant BldR2 as determined by MS analysis was 16348 Da, in agreement with the corresponding theoretical value, and the same applies to the mutants. The quaternary structure was assessed via analytical gel filtration and revealed a homodimeric structure both at different protein concentrations and in different buffers (Figure 2B and Figure S2 of the Supporting Information), a result consistent with other MarR homologues.¹⁵

Specific Binding of BldR2 to Its Promoter Region. Because the majority of MarR transcription factors are autoregulators that bind site-specifically to their promoters, we assessed whether BldR2 had such a capacity. An EMSA confirmed that BldR2 was able to bind to this region (Figure 2C, lane 2) and revealed the specificity of the interaction; in fact, a *gfp* gene fragment²⁶ at a 250-fold molar excess could not compete for BldR2 binding (Figure 2C, lane 4), whereas an unlabeled specific competitor abolished gel retardation when added at the same ratio (Figure 2C, lane 3). Hence, this result shows that BldR2 specifically recognizes its own promoter.

Structural and Functional Characterization of BldR2 and Its Mutants. A multiple-sequence alignment of BldR2 with archaeal MarR members whose structure is known and comparison with bacterial representatives identified four identical residues, namely, Leu29, Leu74, Thr97, and Gly100 (Figure 3A). Leu29, Thr97, and Gly100 are located in helices α 1 and α 5 in the dimerization domain, while Leu74 is in helix α 4 in the DNA binding domain.^{13,24} Residues Glu75, Arg89, and Glu95,

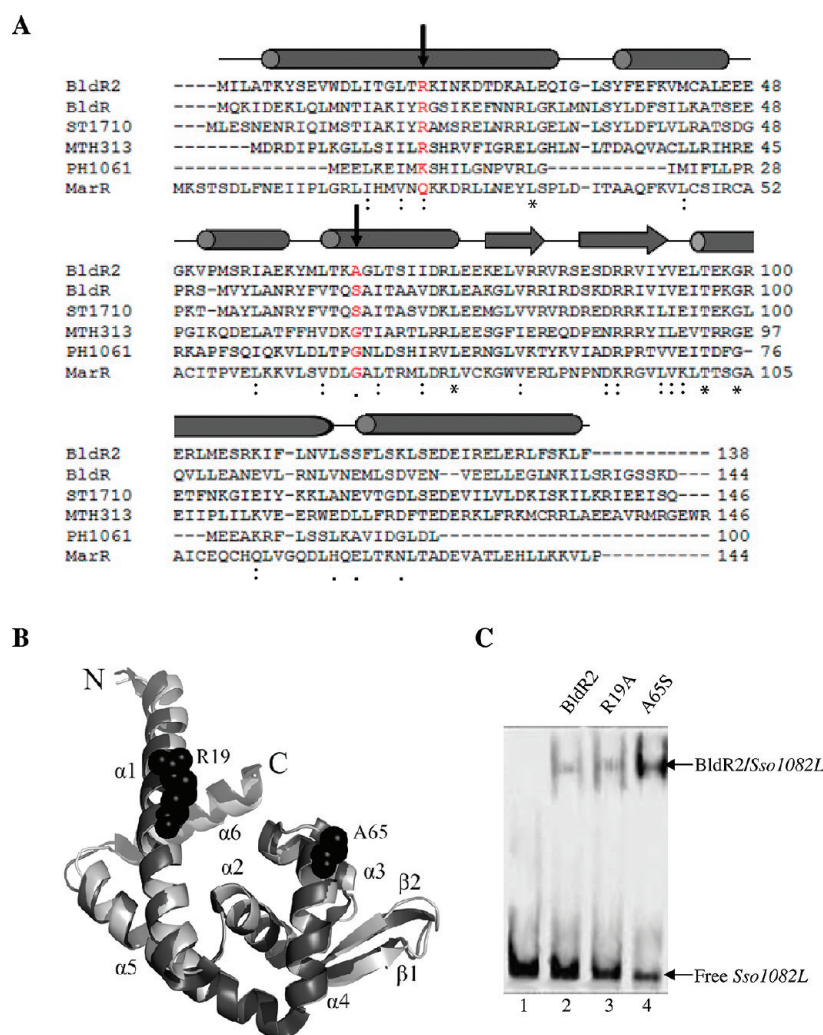


Figure 3. (A) Multiple-sequence alignment of BldR2 (*Sso1082*) with characterized MarR members. Proteins are BldR (*Sso1352*) from *S. solfataricus*, ST1710 from *S. tokodaii*, MTH313 from *M. thermoautotrophicum*, PH1061 from *P. horikoshii*, and MarR from *E. coli*. The secondary structure elements of BldR2 are depicted above the sequences. The mutations introduced by site-directed mutagenesis are highlighted by arrows. (B) Superposition of the 3D structure of the monomer of BldR (dark gray, Protein Data Bank entry 3F3X) and the 3D model of BldR2 (light gray). The mutated residues are shown as gray spheres. (C) Binding of purified BldR2, R19A, and A65S to the *Sso1082L* promoter region: lane 1, *Sso1082L*; lane 2, *Sso1082L* and 15 μ M BldR2; lane 3, *Sso1082L* and 15 μ M R19A; lane 4, *Sso1082L* and 15 μ M A65S.

occurring in strand β 3 (the wing motif), are identical only in the archaeal domain. Hence, on the basis of sequence analysis, BldR2 is expected to share a function similar to that of the other family members.

Sequence analysis of BldR2 from *S. solfataricus*, in comparison with its closest homologue, BldR, from the same organism, provided further information for the design of mutants that could help shed light on the structural properties and functions of BldR2. On the basis of the homology sequence shared between BldR and BldR2, we used the known 3D structure of BldR as a reference model for BldR2 (Figure 3B).²⁴ By also looking at the sequence alignment of Figure 3A, we found that some of the key residues important for dimer stabilization and DNA interaction in BldR are present in BldR2. Among these, the arginine in position 19 has been found to be conserved in known archaeal representatives. In most members with known structure, the Arg residue is in helix α 1 of the dimerization domain and is located at the dimer interface where it contributes to dimer stability.^{15,24} Furthermore, BldR2 has an alanine in position 65 instead of a

serine that is conserved in BldR and ST1710 and is involved in DNA binding.^{17,24} To investigate the contribution of these two amino acids to the DNA binding properties as well as the stability of BldR2, two mutants were generated. In the first, the arginine at position 19 was substituted with an alanine; in the second, the alanine at position 65 was substituted with a serine. For each of the single mutants, we performed a characterization in parallel with the wild-type protein. The mutants were overexpressed in *E. coli* and purified using the same procedure described for the wild-type enzyme. Native gel filtration revealed their ability to form dimers, suggesting that the mutations introduced did not alter the monomer–monomer interactions (Figure S2 of the Supporting Information). To verify whether the mutations affected the DNA binding capability of BldR2, we employed EMSAs using the *Sso1082L* promoter as the target DNA. Figure 3C shows that both mutants retained their activity, again indicating that the mutations were not disruptive. Interestingly, an increased intensity of the shifted band could be observed when the promoter region was incubated with identical amounts

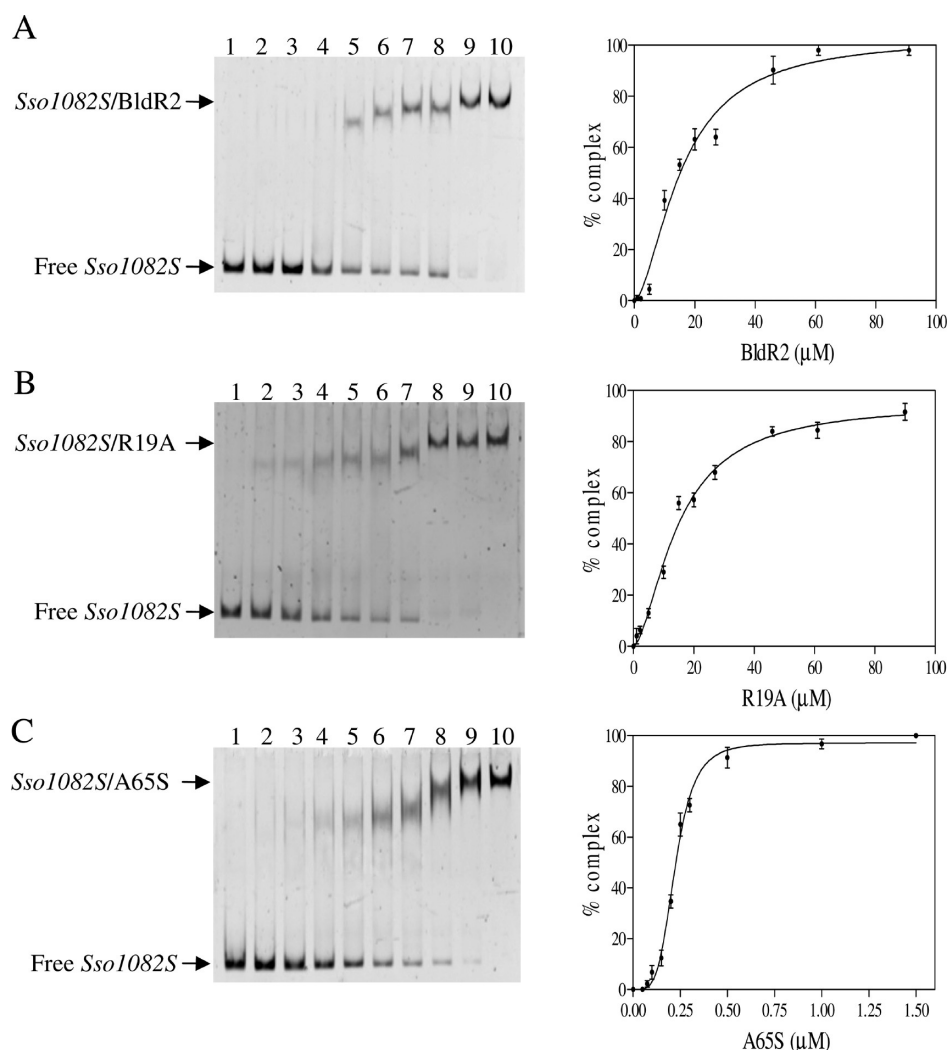


Figure 4. Binding of BldR2, A65S, and R19A to the *Sso1082* promoter assessed by an EMSA. (A) *Sso1082S* (150 nM) titrated with increasing concentrations of BldR2: lane 1, labeled DNA fragment; lanes 2–10, DNA probe incubated with BldR2 at concentrations ranging from 1.5 to 91 μM . (B) *Sso1082S* (150 nM) titrated with increasing concentrations of R19A: lane 1, labeled DNA fragment; lanes 2–9, *Sso1082S* incubated with R19A at concentrations ranging from 1.5 to 91 μM . (C) *Sso1082S* (150 nM) titrated with increasing concentrations of A65S: lane 1, labeled DNA fragment; lanes 2–9, *Sso1082pr* incubated with A65S at concentrations ranging from 0.05 to 1.5 μM . Densitometric data from EMSA obtained as described in Materials and Methods are plotted vs the concentration of each protein (right of each panel). Error bars represent the standard deviation for each point derived from four experiments.

of the A65S mutant (Figure 3C, lane 4) with respect to the wild-type protein (lane 2) and R19A (lane 3).

DNA Interaction Studies. To gain insight into the biological role of BldR2, we analyzed in more detail its DNA binding compared to that of its mutants, testing by an EMSA their binding affinity for an *Sso1082* promoter region, spanning positions –184 to +34 relative to the transcription and translation initiation codon (*Sso1082S*).

Titration with increasing amounts of the BldR2 dimer indicated that the protein binds its own control region in a concentration-dependent manner; furthermore, at saturating concentrations, the protein determined a shift with decreased mobility, suggesting either that more binding sites with different affinities could exist in the DNA sequence analyzed or that multiple dimers could associate to the cognate DNA. The profile obtained by fitting densitometric data to a binding curve with a Hill slope gave an overall apparent equilibrium dissociation

Table 1. Dissociation Constants Calculated by EMSA for BldR2, R19A, and A65S with the *Sso1082* Promoter

	BldR2	R19A	A65S
<i>Sso1082S</i>	$15.8 \pm 5.2 \mu\text{M}$	$16.3 \pm 4.5 \mu\text{M}$	$0.22 \pm 0.02 \mu\text{M}$

constant (K_d) of 15.8 μM (Figure 4A and Table 1) and provided a Hill coefficient of 1.7. A comparable binding pattern (Figure 4B) with a similar global affinity ($K_d = 16.3 \mu\text{M}$) and slope (1.6) was observed for the R19A mutant, while mutation of A65 to serine significantly changed the DNA binding properties; in fact, formation of a complex with A65S was seen using much lower protein concentrations (Figure 4C), and fitted binding data indicated an ~ 70 -fold increased affinity for the target promoter ($K_d = 0.22 \mu\text{M}$) and a slope of 4.3. These results clearly suggested that A65S binds with high affinity and via a different mechanism with respect to those of BldR2 and R19A.

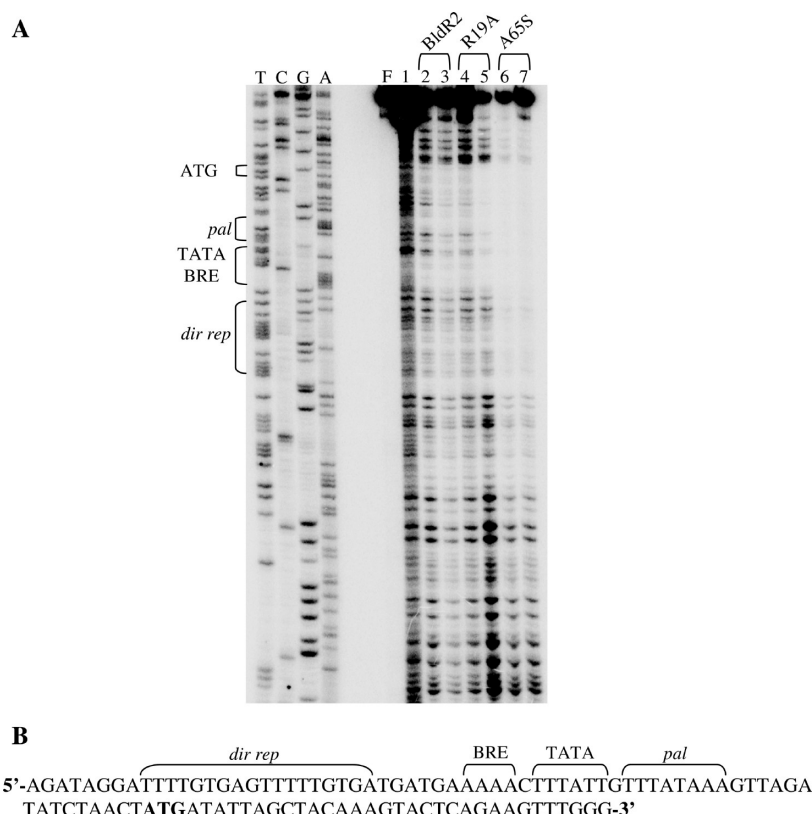


Figure 5. (A) Wild-type and mutant binding sites of the *Sso1082* promoter. DNase I footprinting analysis was performed at the nontemplate strand using 0.0 μ g (lane 1), 5 μ g ($\sim 15 \mu$ M, lanes 2, 4, and 6), and 10 μ g (lanes 3, 5, and 7) of purified BldR2, R19A, and A65S. DNA fragments were analyzed in parallel with a sequencing reaction (relative lanes are indicated by the corresponding nucleotide positions on the top) by denaturing gel electrophoresis. (B) Positions of the footprints indicated on the nucleotide sequences relative to the transcription and translation start site.

With the aim of precisely positioning the binding sites of BldR2 at its own promoter, a DNase I footprinting analysis was undertaken. As shown in Figure 5A, BldR2, R19A, and A65S protect a region of ~ 40 bp extending from the pseudopalindromic TATA-BRE sequences to the ATG start codon and containing a 8 bp TTTATAAA palindromic site (Figure 5B). The extent of protection suggests that more than one BldR2 dimer could associate to target DNA. Furthermore, these results prove the presence of regulatory sites in the BldR2 promoter that could function for its *in vivo* autoregulation. Interestingly, the A65S mutant gave an even more extended footprint; in fact, it protected a further region of 17 bp located more upstream of the TATA box, containing the direct repeat TTTTGTGAgTTTGTGA (Figure 5B). This evidence could imply that the introduction of a serine enhances the DNA binding affinity by modifying the recognition sequence.

To analyze the DNA binding behavior of BldR2 upon addition of putative phenolic ligands, we performed EMSAs in which formation of the complex of BldR2 and the mutants with the *Sso1082S* promoter was tested after preincubation of the proteins with increasing concentrations of salicylate and benzaldehyde. The first ligand was chosen because it was demonstrated to be the negative effector of several MarR homologues,^{17,22,41} while the second was known to interact positively with the BldR factor in *S. solfataricus*.²⁵

The results are shown in Figure 6A: an inhibition of complex formation starts at ~ 10 mM salicylate (lane 5), while the benzaldehyde was slightly less effective (Figure 6B). In fact, release of

BldR2 from its own promoter could be observed from ~ 30 mM benzaldehyde (lane 6).

DNA binding by the mutant A65S was also affected in the presence of the tested ligands (Figure 6C,D) to a similar extent; in addition, we observed a further band above that of the free DNA, which was interpreted as a partial dissociation of the DNA–protein complex upon ligand interaction.

Taken together, these data indicate that salicylate and benzaldehyde are low-affinity ligands of BldR2 that associate with the protein to attenuate DNA binding.

Conformational Stability of Dimeric BldR2. According to its hyperthermophilic origin, BldR2 has proved to be highly thermostable and resistant to the GuHCl denaturing action as supported by detailed investigation of its conformational stability by means of CD and fluorescence measurements. Figure 7 shows the CD spectra of BldR2 in the far-UV region at 25 °C (spectrum a) and 105 °C (spectrum b). From the latter spectrum, it appears that BldR2 is still folded at that high temperature; in fact, it retains the characteristic minima at 222 and 208 nm with a small decrease in the CD band intensities. The complete disappearance of the canonical CD bands occurs when the protein is incubated for 24 h with 7 M GuHCl (spectrum c). The analysis of the CD spectrum at 25 °C, performed using Dichroweb,³⁷ suggested that BldR2 contains 66% α -helix and 9% β -sheet; these values closely resemble those obtained from the X-ray structure of the homologous protein, BldR.^{24,25} The near-UV CD spectrum of BldR2 is reported in the inset of Figure 7. It also suggested that the protein possesses a well-defined conformation at 25 °C in buffer

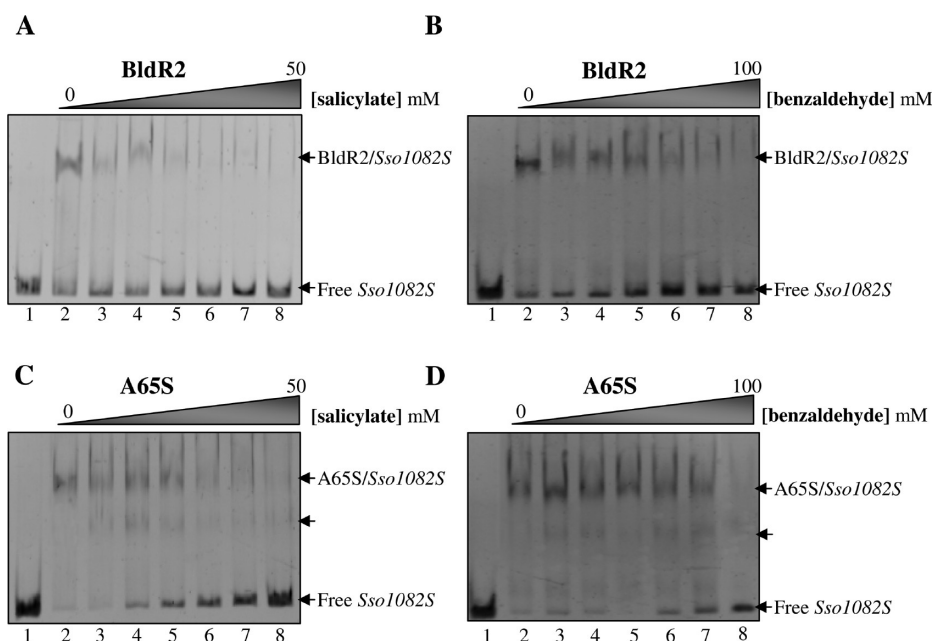


Figure 6. Formation of BldR2–*Sso1082S* and A65S–*Sso1082S* complexes in the presence of salicylate and benzaldehyde. (A and B) *Sso1082S* (150 nM) and BldR2 (16 μ M) titrated with increasing concentrations of salicylate (A) and benzaldehyde (B): lane 1, DNA fragment; lane 2, *Sso1082S* incubated with BldR2; lanes 3–8, *Sso1082S* and BldR2 incubated with 3, 5, 10, 20, 30, and 50 mM salicylate (A) and 3, 10, 20, 30, 50, and 100 mM benzaldehyde (B), respectively. (C and D) *Sso1082S* (150 nM) and A65S (0.2 μ M) titrated with increasing concentrations of salicylate (C) and benzaldehyde (D): lane 1, DNA fragment; lane 2, *Sso1082S* incubated with A65S; lanes 3–8, *Sso1082S* and A65S incubated with 3, 5, 10, 20, 30, and 50 mM salicylate (C) and 3, 10, 20, 30, 50, and 100 mM benzaldehyde (D), respectively.

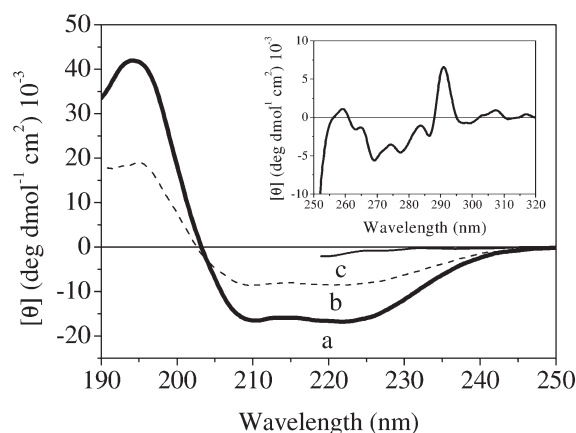


Figure 7. (A) Far-UV CD spectra of BldR2. Spectra were recorded at 25 °C (a, —), 105 °C (b, ---), and 25 °C after incubation for 24 h with 7 M GuHCl (c, —). The inset shows the near-UV CD spectrum of BldR2 (—). The protein concentrations used to acquire the spectra are 2.4 and 20 μ M in the far- and near-UV regions, respectively.

solution. The CD spectra of both mutants A65S and R19A were found to be very similar to those of BldR2, suggesting that the secondary and tertiary structure are not affected by the mutations (Figure S3 of the Supporting Information).

Thermal Unfolding. Changes in the far-UV CD signal at 222 nm have been used to follow the thermal unfolding of BldR2 in 20 mM sodium phosphate buffer (pH 7.5). Because of the very high thermal stability, it was not possible to obtain a complete thermal unfolding curve as the transition is not yet finished at 105 °C [see the filled squares (■) in Figure 8], and the instrumental setup cannot work at

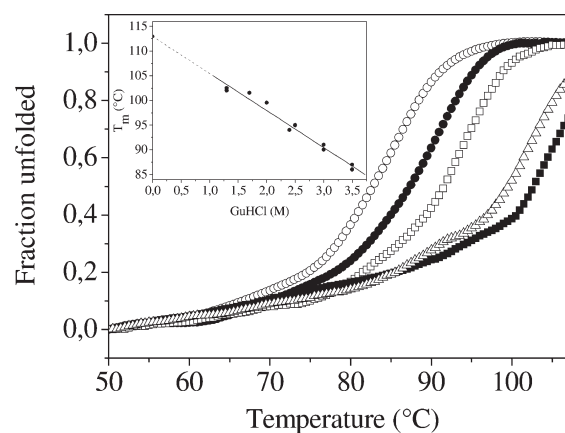


Figure 8. Thermal denaturation curves of BldR2 at a fixed concentration of 2.4 μ M, in 20 mM sodium phosphate (pH 7.5) in the presence of 0 (■), 1.3 (△), 2.4 (□), 3 (●) and 3.5 M GuHCl (○). The curves were obtained by recording the changes in the molar ellipticity at 222 nm as a function of temperature. The inset shows the dependence of the melting temperature T_m on GuHCl concentration; via linear extrapolation, a value for T_m in the absence of the denaturant agent was estimated.

temperatures higher than 110 °C. To obtain complete thermal denaturation curves, solutions of BldR2 at a fixed final concentration of 2.4 μ M were incubated with increasing GuHCl concentrations to progressively destabilize the native state. The corresponding thermal denaturation curves are collected in Figure 8, and the denaturation temperatures obtained were as follows: 102 °C at 1.3 M GuHCl, 96 °C at 2.4 M GuHCl, 91 °C at 3 M GuHCl, and 85 °C at 3.5 M GuHCl. Via linear extrapolation of these numbers up to 0 M GuHCl, the estimated denaturation temperature of BldR2

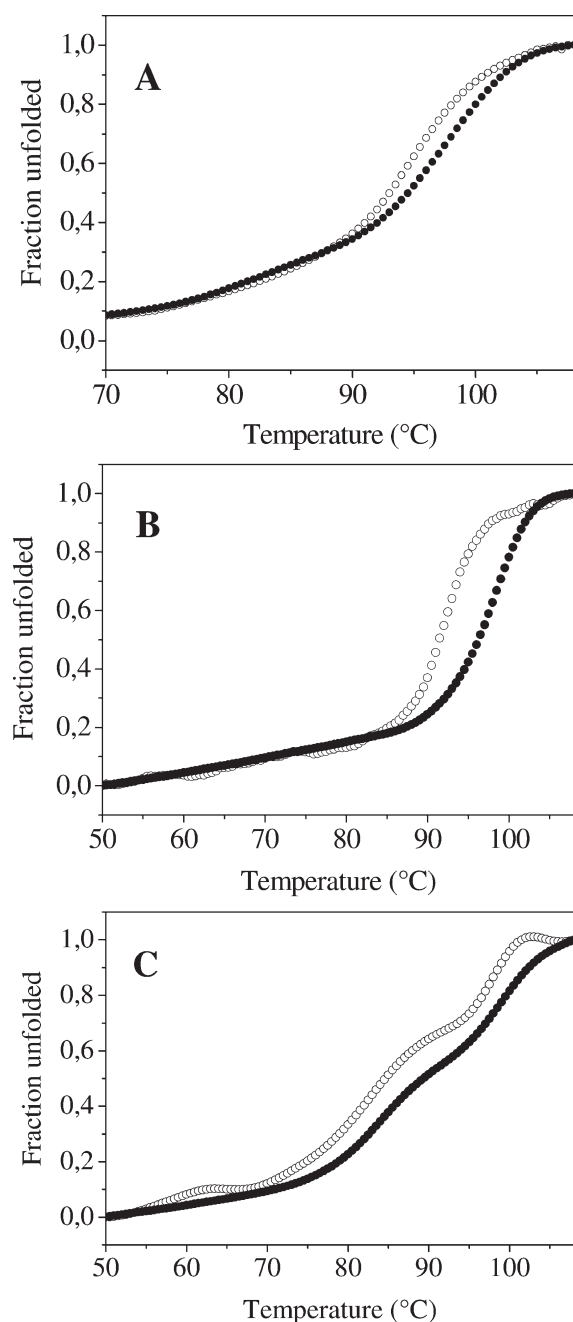


Figure 9. (A) Concentration dependence of BldR2 thermal denaturation, in the presence of 2.4 M GuHCl. Protein concentrations were 1.5 (○) and 9 μM (●), and the T_m shifted from 94.5 to 97.5 °C. (B) Concentration dependence of A65S thermal denaturation, in the presence of 2.4 M GuHCl. Protein concentrations were 0.9 (○) and 9 μM (●), and the T_m shifted from 92.5 to 98 °C. (C) Concentration dependence of R19A thermal denaturation, in the presence of 2.4 M GuHCl. Protein concentrations were 0.9 (○) and 9 μM (●).

is 113 ± 1 °C (inset of Figure 8). Figure 9 shows the dependence of the denaturation temperature of BldR2 and its mutants, A65S and R19A, in the presence of 2.4 M GuHCl at two different protein concentrations. For BldR2 and A65S, we observed a shift in the denaturation temperature in a single sigmoid-shaped curve indicative of a two-state dimeric unfolding process (Figure 9A,B), whereas thermal denaturation curves of R19A confirm the concentration

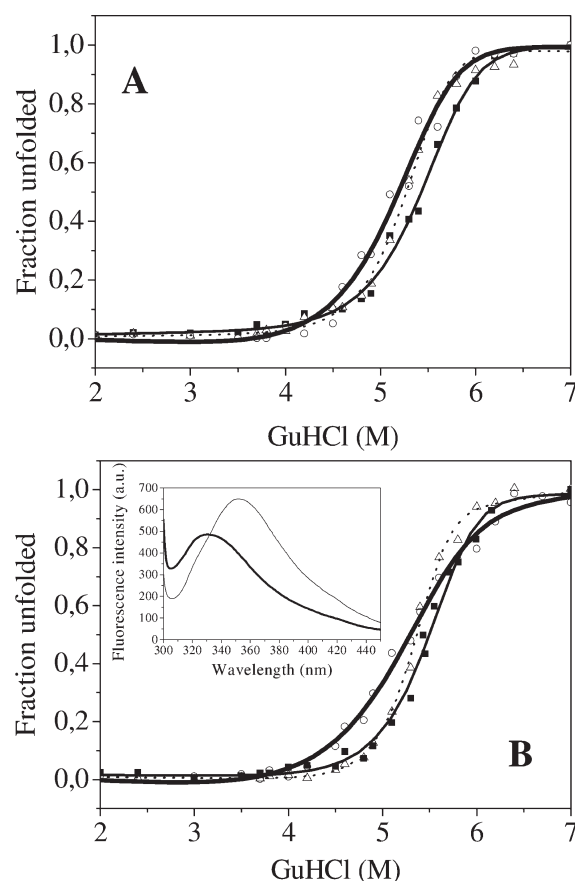


Figure 10. (A) GuHCl-induced unfolding curves of BldR2 (■), A65S (△), and R19A (○), obtained by recording the molar ellipticity at 230 nm and 25 °C. (B) Change in the fluorescence intensity at 336 nm and 25 °C. The lines are the best fits of the curves performed as described in Materials and Methods (inset of Figure 5B). Fluorescence emission spectra of native BldR2 in buffer solution (pH 7.5) (○) and in the presence of 7 M GuHCl (●) at 25 °C.

dependence but show a biphasic sigmoidal shape with two inflection points (Figure 9C). The thermal denaturation temperature of A65S proved to be very similar to that of BldR2 (111 ± 1 °C). Instead, the biphasic curve of R19A prevented us from correctly estimating the denaturation temperature and strongly suggested a more complex thermal unfolding process (Figure S4 of the Supporting Information). Because thermal denaturation of BldR2 and the mutants is not fully reversible under any of the investigated experimental conditions, no further thermodynamic analysis was performed.

Denaturant-Induced Unfolding. The conformational stability of BldR2 and its mutants against the denaturing action of GuHCl has been investigated at 25 °C in 20 mM sodium phosphate buffer (pH 7.5) by performing CD and fluorescence measurements. It is worth noting that urea cannot be used because BldR2 is extremely resistant to this denaturing agent. The transition curves obtained by recording the molar ellipticity at 230 nm (i.e., detecting the secondary structure stability) and that obtained by recording the fluorescence intensity (tertiary structure) are reported in panels A and B of Figure 10. The inset of Figure 10B shows the comparison of fluorescence emission spectra of BldR2 in the absence and presence of 7 M GuHCl when the protein is completely unfolded according to the

Table 2. Thermodynamic Parameters Obtained by the Analysis of the GuHCl-Induced Unfolding Curves of BldR2 Monitored by the Change in Molar Ellipticity at 230 nm ($[\theta]_{230}$) and of Fluorescence Intensity I_{336} at pH 7.5, 20 mM Sodium Phosphate Buffer, and 25 °C^a

sample	probe	$\Delta G_U(\text{H}_2\text{O})$ (kJ/mol)	m (kJ mol ⁻¹ M ⁻¹)	C_m (M)
BldR2	$[\theta]_{230}$	101 ± 5	13 ± 1	5.6
	I_{336}	106 ± 5	14 ± 1	5.6
A65S	$[\theta]_{230}$	116 ± 3	16 ± 1	5.4
	I_{336}	119 ± 4	16 ± 1	5.4
R19A	$[\theta]_{230}$	87 ± 6	11 ± 1	5.2
	I_{336}	82 ± 3	10 ± 1	5.2

^a Each figure is the average of the values calculated by the nonlinear regression procedure over three independent measurements. Errors are the standard deviations of the fits.

exposure of Trp and Tyr residues to the aqueous solvent. The different spectroscopic probes show superposable transition curves, indicating that both the secondary and tertiary structures are concurrently lost. The GuHCl-induced denaturation curves have a simple sigmoid shape for BldR2, A65S, and R19A, suggesting a cooperative two-state transition between folded dimers and unfolded monomers. The GuHCl-induced unfolding has proven to be reversible: fully unfolded samples showed recovery of spectroscopic features of the native protein after suitable dilution. The GuHCl concentrations at half-completion of the transition are 5.6, 5.4, and 5.2 M for BldR2, A65S, and R19A, respectively, highlighting the very high resistance to the denaturant. As the GuHCl-induced unfolding of BldR2 and A65S and R19A mutants is a reversible process, a thermodynamic analysis of the transitions can be performed. Linear extrapolation of ΔG_U (eq 3) yields the unfolding Gibbs energy in the absence of denaturant [$\Delta G_U(\text{H}_2\text{O})$] and the m value; all of the thermodynamic data of the unfolding transitions are listed in Table 2.

DISCUSSION

Because of their function in multidrug resistance and tolerance to highly toxic compounds, MarR transcriptional regulators have been intensely characterized in bacteria and archaea with the aim of understanding the molecular mechanisms of regulated response to such a stress. As many archaea live in hostile environments and are able to defend themselves from a wide variety of stress conditions, they represent an interesting model for determining their ability to survive under rapidly changing conditions. For such organisms, MarR-type regulators might be critical for their adaptation to particular habitats or lifestyles. Intriguingly, the exploration of homologous protein sequences from *Sulfolobales*, including different strains of *S. solfataricus*, *Sulfolobus islandicus*, *S. tokodaii*, and *Sulfolobus acidocaldarius*, revealed that, differently from the majority of bacteria, they all have at least two conserved MarR representatives, indicating the presence of several still uncharacterized regulatory systems involved in multiple-antibiotic resistance. MarR proteins also represent outstanding models for the study of protein stability for several reasons. (i) Many structures from distantly related organisms are available. (ii) They have a dimeric quaternary structure. (iii) They are able to bind DNA only as homodimers.

In this study, an archaeal MarR member, BldR2, was overexpressed in *E. coli* and purified to homogeneity. The protein forms a homodimer in solution and binds specifically to its own

promoter region with micromolar affinity, a value that is comparable to those other archaeal MarR members^{17,24,25} but is lower than those of many bacterial counterparts.

A primer extension analysis identified a transcription initiation site mapping 45 bp downstream of the first computationally predicted *Sso1082* codon³⁰ and according to transcriptome mapping.³¹ On the basis of our result, a promoter region was found containing consensus BRE-TATA sequences centered at positions −31 and −27 with respect to the transcription start site.

The binding site for BldR2 as determined by DNase I footprinting analysis was found in a region extending from BRE/TATA sequence to the transcription start site. Interestingly, the bound region contains an 8 bp pseudopalindromic sequence (BRE/TATA, AAACCTTTA) separated by 3 bp from an 8 bp perfect palindrome (TTTATAAA), suggesting that one BldR2 dimer could bind to each site. Hence, according to the location of the identified basal promoter elements and the BldR2 binding region, it can be proposed that the protein interferes with *Sso1082* transcription by competing with basal transcription factors.

Benzaldehyde and salicylate, known to act as effectors of different MarR proteins at millimolar concentrations, were able to weaken the interaction of BldR2 with its own promoter, suggesting that salicylate and to a lesser extent benzaldehyde could be ligands for BldR2. The low affinity of the effectors for the protein raises questions about their physiological relevance and indicates that other aromatic compounds could be the natural effectors. The in vitro binding results correlated with the in vivo induction of *BldR2* gene expression upon addition of aromatic drugs; the level of gene expression was also increased during the late-log growth phase. The derepression of *BldR2* both in the presence of aromatic compounds and in late-log growth phase supports a picture in which *BldR2* expression could be regulated by endogenous effectors derived from aromatic catabolic pathways. Hence, BldR2 in vivo would control regulatory mechanisms diverse from those regulated by BldR, which mainly works in the exponential growth phase, and in a different way.²⁵

Two single-point mutants, R19A and A65S, were also produced and characterized. The dimeric state of the mutants was confirmed by gel filtration experiments and thermal unfolding at different protein concentrations. The results of the thermodynamic characterization showed that BldR2 possesses a very stable dimeric conformation and that the mutants have also a very high resistance to both temperature and GuHCl denaturing action. In fact, the estimated denaturation temperature of BldR2 is in good agreement with that of the homologous ST1710 from *S. tokodaii*, obtained by differential scanning calorimetry¹⁷ and of the dimeric protein ORF56 from *S. islandicus*.⁴⁰ The monophasic temperature-induced unfolding curve of the wild-type protein suggested that the denaturation mechanism occurs in the absence of detectable equilibrium intermediates. This behavior is conserved in mutant A65S, suggesting that the mutation does not affect the protein global stability in solution. Interestingly, mutant R19A shows a biphasic sigmoid-shaped thermally induced unfolding curve. This finding suggests that, in this case, a more complex process occurs, going through the formation of stable intermediate species. On the basis of those results, it is tempting to speculate that residue R19 could be involved in important stabilizing interactions in the dimerization region of the protein, in a way that is reminiscent of what happens in the BldR protein. In fact, in BldR the conserved Arg19 residue of one monomer sets up a strong H bond (distance of 3 Å) with Tyr60 of the other

monomer.²⁴ The temperature-induced unfolding process is not reversible for all of the studied protein samples; this thermal irreversibility seems to be a common feature of MarR family members. In fact it was also found for the mesophilic HucR from *Deinococcus radiodurans*.⁴¹

The GuHCl-induced unfolding transitions of BldR2, monitored by CD in the far-UV region and fluorescence, are equivalent, indicating that both the protein secondary and tertiary structure are lost at the same time. This allows us to argue that the transition can be described by a two-state model for a dimer. The experimental values of C_m for BldR2 (5.6 M) and mutants A6SS (5.4 M) and R19A (5.2 M) indicate that the wild type is more resistant to the denaturant action than the mutants. The Gibbs energy of unfolding (101 kJ mol^{-1}) is in agreement with that of ORF56 from *S. islandicus* (85.1 kJ mol^{-1}).^{40,42} A closer analysis of ΔG_U and m values yields intriguing information about the structural properties of the wild-type protein and its mutants. In particular, even if A6SS has a C_m value lower than that of BldR2, its m value is higher and, as a consequence, ΔG_U values are very close. Because a higher m value is usually related to a larger change in the accessible surface area (ASA) upon unfolding, we can argue that the unfolded conformation of the A6SS mutant may be more exposed to the solvent.

Another interesting finding is represented by the denaturant-induced unfolding results of R19A that clearly show a simple sigmoid-shaped curve quite different from the biphasic profile obtained for the temperature unfolding curves. The result of the interpolation procedure, based on a simple two-state model, gives an overall lower stability of R19A compared to that of the wild-type protein; it is possible that this value represents an underestimation mainly caused by the choice of the model. In fact, it has been reported previously⁴³ that the presence of intermediate species in the chemically induced unfolding could not produce any dramatic perturbation in the monophasic unfolding curve, but in that case, the assumption of a two-state model would underestimate the Gibbs energy and the m value.⁴³ This conclusion fully agrees with the scenario that we previously anticipated on the basis of the thermal experiments in which Arg19 represents an important player in the stabilization of the dimer. Taken together, these results raise the possibility that a simple two-state model could not fully describe the BldR2 denaturation process. Possibly, the equilibrium unfolding of the BldR2 dimer could be described as an apparent two-state reversible reaction, in which unfolding and dissociation are coupled processes. This apparent two-state reaction seems to be confirmed by A6SS behavior, while for R19A, the unfolding and dissociation processes seem to be separated, because of the key position of the arginine residue in the dimerization domain.

The structure of the archaeal ST1710–DNA complex and a molecular model of the BldR–DNA complex showed that serine 65 was, in both cases, a critical residue for protein–DNA interaction acting as both a donor and an acceptor of H-bonds with its target DNA. Interestingly, the Ser68 in the bacterial OhrR of *Bacillus subtilis* was also found to make contacts with the major groove of the DNA molecule.⁴⁴ In our study, the characterization of the DNA binding properties of the A6SS mutant highlighted Ser65 as a key amino acid; in fact, this single-amino acid substitution was able not only to increase the extent of protein–DNA interaction by ~ 70 -fold but also to cause its binding to a further sequence that is a direct repeat located immediately upstream the TATA/BRE sequence in the *Sso1082*

promoter. In archaea, binding by transcriptional regulators to sequences in that position has been proven to be associated with a transcriptional activation.^{25,45}

This evidence suggests that the serine residue would allow the formation of an additional interaction with DNA responsible for an extension of the recognition sequence; this would cause a modification both in the binding mechanism and in the sequence specificity. On the basis of this observation, we can also hypothesize that this difference could correlate with the different proposed *in vivo* physiological roles of the two BldR proteins and depict A6SS as a protein intermediate between a repressor (BldR2) and an activator (BldR).

In conclusion, we propose for BldR2 a role in its autoregulation, but further analyses are required to understand the overall mechanism of regulation by BldR2, which would include a deeper investigation of the *in vivo* function of multiple transcription start sites, a genome-wide identification of target genes and natural ligands, and the analysis of BldR[−] and BldR2[−] mutant cells. Furthermore, this study highlights the idea that the identification of key residues involved in dimer stability may contribute to our understanding of the structural–functional relationship in the MarR family, because it is known that mutations in the dimerization domain are critical for the transcriptional regulation of MarR members.¹⁹ Moreover, given that mutations in the DNA binding domain can increase antibiotic tolerance,⁴⁶ we also suggest that knowledge of amino acids involved in DNA recognition may provide a remarkable starting point for the design of engineered MarR regulators acting as innovative therapeutic tools.

■ ASSOCIATED CONTENT

S Supporting Information. A densitometric analysis of a Northern blot, gel filtration profiles, far-UV CD spectra, and thermal denaturation curves at different GuHCl concentrations of A6SS and R19A. This material is available free of charge via the Internet at <http://pubs.acs.org>.

■ AUTHOR INFORMATION

Corresponding Author

*Phone: +39081679167 (G.F.) or +39081674255 (P.D.V.). Fax: +39081679053. E-mail: fiogabri@unina.it (G.F.) or pompea.delvecchio@unina.it (P.D.V.).

Funding Sources

This work was supported by MIUR-PRIN (Ministero dell'Istruzione, dell'Università e della Ricerca Scientifica-Progetti di Ricerca di Interesse Nazionale) (Grant 2008, CUP: E61J100000200001). The Centro Interdipartimentale di Metodologie Chimico Fisiche (CIMCF, University of Naples Federico II) is gratefully acknowledged for providing the spectrometers. L.M. is a fellow of the European Molecular Biology Organization (EMBO).

■ ACKNOWLEDGMENT

We acknowledge Dr. Raffaele Ronca for helpful discussion, Dr. Patrizia Contursi for critically reading the manuscript, and Dr. Omri Wurtzel for precious explanations about *S. solfataricus* transcriptomic data.

■ ABBREVIATIONS

BldR2, MarR member from *S. solfataricus*; IPTG, isopropyl 1-thio- β -D-galactopyranoside; SDS–PAGE, sodium dodecyl sulfate–polyacrylamide gel electrophoresis; EMSA, electrophoretic mobility shift assay; GuHCl, guanidine hydrochloride; CD, circular dichroism.

■ REFERENCES

- (1) Wheelis, M. L., Kandler, O., and Woese, C. R. (1992) On the nature of global classification. *Proc. Natl. Acad. Sci. U.S.A.* 89, 2930–2934.
- (2) Grabowski, B., and Kelman, Z. (2003) Archeal DNA replication: Eukaryal proteins in a bacterial context. *Annu. Rev. Microbiol.* 57, 487–516.
- (3) Verhees, C. H., Kengen, S. W., Tuininga, J. E., Schut, G. J., Adams, M. W., De Vos, W. M., and Van Der Oost, J. (2003) The unique features of glycolytic pathways in Archaea. *Biochem. J.* 375, 231–246.
- (4) van de Vossenberg, J. L., Driessen, A. J., and Konings, W. N. (1998) The essence of being extremophilic: The role of the unique archaeal membrane lipids. *Extremophiles* 2, 163–170.
- (5) Geiduschek, E. P., and Ouhammouch, M. (2005) Archaeal transcription and its regulators. *Mol. Microbiol.* 56, 1397–1407.
- (6) Hirata, A., Klein, B. J., and Murakami, K. S. (2008) The X-ray crystal structure of RNA polymerase from Archaea. *Nature* 451, 851–854.
- (7) Ramos, A., Raven, N., Sharp, R. J., Bartolucci, S., Rossi, M., Cannio, R., Lebbink, J., Van Der Oost, J., De Vos, W. M., and Santos, H. (1997) Stabilization of Enzymes against Thermal Stress and Freeze-Drying by Mannosylglycerate. *Appl. Environ. Microbiol.* 63, 4020–4025.
- (8) Jaenicke, R., and Zavodszky, P. (1990) Proteins under extreme physical conditions. *FEBS Lett.* 268, 344–349.
- (9) Razvi, A., and Scholtz, J. M. (2006) Lessons in stability from thermophilic proteins. *Protein Sci.* 15, 1569–1578.
- (10) Pedone, E., Bartolucci, S., and Fiorentino, G. (2004) Sensing and adapting to environmental stress: The archaeal tactic. *Front. Biosci.* 9, 2909–2926.
- (11) Schelet, J., Drozda, M., Dixit, V., Dillman, A., and Blum, P. (2006) Regulation of mercury resistance in the crenarchaeote *Sulfolobus solfataricus*. *J. Bacteriol.* 188, 7141–7150.
- (12) Limauro, D., Pedone, E., Pirone, L., and Bartolucci, S. (2006) Identification and characterization of 1-Cys peroxiredoxin from *Sulfolobus solfataricus* and its involvement in the response to oxidative stress. *FEBS J.* 273, 721–731.
- (13) Wilkinson, S. P., and Grove, A. (2006) Ligand-responsive transcriptional regulation by members of the MarR family of winged helix proteins. *Curr. Issues Mol. Biol.* 8, 51–62.
- (14) Newberry, K. J., Fuangthong, M., Panmanee, W., Mongkolsuk, S., and Brennan, R. G. (2007) Structural mechanism of organic hydroperoxide induction of the transcription regulator OhrR. *Mol. Cell* 28, 652–664.
- (15) Alekshun, M. N., Levy, S. B., Mealy, T. R., Seaton, B. A., and Head, J. F. (2001) The crystal structure of MarR, a regulator of multiple antibiotic resistance, at 2.3 Å resolution. *Nat. Struct. Biol.* 8, 710–714.
- (16) Perera, I. C., and Grove, A. (2010) Molecular mechanisms of ligand-mediated attenuation of DNA binding by MarR family transcriptional regulators. *J. Mol. Cell Biol.* 2, 243–254.
- (17) Kumarevel, T., Tanaka, T., Umehara, T., and Yokoyama, S. (2009) ST1710-DNA complex crystal structure reveals the DNA binding mechanism of the MarR family of regulators. *Nucleic Acids Res.* 37, 4723–4735.
- (18) Perera, I. C., Lee, Y. H., Wilkinson, S. P., and Grove, A. (2009) Mechanism for attenuation of DNA binding by MarR family transcriptional regulators by small molecule ligands. *J. Mol. Biol.* 390, 1019–1029.
- (19) Okada, N., Oi, Y., Takeda-Shitaka, M., Kanou, K., Umeyama, H., Haneda, T., Miki, T., Hosoya, S., and Danbara, H. (2007) Identification of amino acid residues of *Salmonella* SlyA that are critical for transcriptional regulation. *Microbiology (Reading, U.K.)* 153, 548–560.

- (20) Miyazono, K., Tsujimura, M., Kawarabayasi, Y., and Tanokura, M. (2007) Crystal structure of an archaeal homologue of multidrug resistance repressor protein, EmrR, from hyperthermophilic archaea *Sulfolobus tokodaii* strain 7. *Proteins* 67, 1138–1146.
- (21) Kumarevel, T., Tanaka, T., Nishio, M., Gopinath, S. C., Takio, K., Shinkai, A., Kumar, P. K., and Yokoyama, S. (2008) Crystal structure of the MarR family regulatory protein, ST1710, from *Sulfolobus tokodaii* strain 7. *J. Struct. Biol.* 161, 9–17.
- (22) Saridakis, V., Shahinas, D., Xu, X., and Christendat, D. (2008) Structural insight on the mechanism of regulation of the MarR family of proteins: High-resolution crystal structure of a transcriptional repressor from *Methanobacterium thermoautotrophicum*. *J. Mol. Biol.* 377, 655–667.
- (23) Okada, U., Sakai, N., Yao, M., Watanabe, N., and Tanaka, I. (2006) Structural analysis of the transcriptional regulator homolog protein from *Pyrococcus horikoshii* OT3. *Proteins* 63, 1084–1086.
- (24) Di Fiore, A., Fiorentino, G., Vitale, R. M., Ronca, R., Amodeo, P., Pedone, C., Bartolucci, S., and De Simone, G. (2009) Structural analysis of BldR from *Sulfolobus solfataricus* provides insights into the molecular basis of transcriptional activation in archaea by MarR family proteins. *J. Mol. Biol.* 388, 559–569.
- (25) Fiorentino, G., Ronca, R., Cannio, R., Rossi, M., and Bartolucci, S. (2007) MarR-like transcriptional regulator involved in detoxification of aromatic compounds in *Sulfolobus solfataricus*. *J. Bacteriol.* 189, 7351–7360.
- (26) Fiorentino, G., Ronca, R., and Bartolucci, S. (2009) A novel *E. coli* biosensor for detecting aromatic aldehydes based on a responsive inducible archaeal promoter fused to the green fluorescent protein. *Appl. Microbiol. Biotechnol.* 82, 67–77.
- (27) Brock, T. D., Brock, K. M., Belly, R. T., and Weiss, R. L. (1972) *Sulfolobus*: A new genus of sulfur-oxidizing bacteria living at low pH and high temperature. *Arch. Mikrobiol.* 84, 54–68.
- (28) Cannio, R., Fiorentino, G., Rossi, M., and Bartolucci, S. (1999) The alcohol dehydrogenase gene: Distribution among *Sulfolobales* and regulation in *Sulfolobus solfataricus*. *FEMS Microbiol. Lett.* 170, 31–39.
- (29) Limauro, D., Falciatore, A., Basso, A. L., Forlani, G., and De Felice, M. (1996) Proline biosynthesis in *Streptococcus thermophilus*: Characterization of the proBA operon and its products. *Microbiology (Reading, U.K.)* 142 (Part 11), 3275–3282.
- (30) She, Q., Singh, R. K., Confalonieri, F., Zivanovic, Y., Allard, G., Awayez, M. J., Chan-Weiher, C. C., Clausen, I. G., Curtis, B. A., De Moors, A., Erauso, G., Fletcher, C., Gordon, P. M., Heikamp-de Jong, I., Jeffries, A. C., Kozera, C. J., Medina, N., Peng, X., Thi-Ngoc, H. P., Redder, P., Schenk, M. E., Theriault, C., Tolstrup, N., Charlebois, R. L., Doolittle, W. F., Duguet, M., Gaasterland, T., Garrett, R. A., Ragan, M. A., Sensen, C. W., and Van der Oost, J. (2001) The complete genome of the crenarchaeon *Sulfolobus solfataricus* P2. *Proc. Natl. Acad. Sci. U.S.A.* 98, 7835–7840.
- (31) Wurtzel, O., Sapra, R., Chen, F., Zhu, Y., Simmons, B. A., and Sorek, R. (2010) A single-base resolution map of an archaeal transcriptome. *Genome Res.* 20, 133–141.
- (32) Notredame, C., Higgins, D. G., and Heringa, J. (2000) T-Coffee: A novel method for fast and accurate multiple sequence alignment. *J. Mol. Biol.* 302, 205–217.
- (33) Lambert, C., Leonard, N., De Bolle, X., and Depiereux, E. (2002) ESyPred3D: Prediction of proteins 3D structures. *Bioinformatics* 18, 1250–1256.
- (34) Bradford, M. M. (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* 72, 248–254.
- (35) Gill, S. C., and von Hippel, P. H. (1989) Calculation of protein extinction coefficients from amino acid sequence data. *Anal. Biochem.* 182, 319–326.
- (36) Compton, L. A., and Johnson, W. C., Jr. (1986) Analysis of protein circular dichroism spectra for secondary structure using a simple matrix multiplication. *Anal. Biochem.* 155, 155–167.
- (37) Whitmore, L., and Wallace, B. A. (2004) DICHROWEB, an online server for protein secondary structure analyses from circular dichroism spectroscopic data. *Nucleic Acids Res.* 32, W668–W673.

- (38) Pace, C. N. (1986) Determination and analysis of urea and guanidine hydrochloride denaturation curves. *Methods Enzymol.* 131, 266–280.
- (39) Tanford, C. (1970) Protein denaturation. C. Theoretical models for the mechanism of denaturation. *Adv. Protein Chem.* 24, 1–95.
- (40) Zeeb, M., Lipps, G., Lilie, H., and Balbach, J. (2004) Folding and association of an extremely stable dimeric protein from *Sulfolobus islandicus*. *J. Mol. Biol.* 336, 227–240.
- (41) Wilkinson, S. P., and Grove, A. (2004) HucR, a novel uric acid-responsive member of the MarR family of transcriptional regulators from *Deinococcus radiodurans*. *J. Biol. Chem.* 279, 51442–51450.
- (42) Luke, K. A., Higgins, C. L., and Wittung-Stafshede, P. (2007) Thermodynamic stability and folding of proteins from hyperthermophilic organisms. *FEBS J.* 274, 4023–4033.
- (43) Soulages, J. L. (1998) Chemical denaturation: Potential impact of undetected intermediates in the free energy of unfolding and m-values obtained from a two-state assumption. *Biophys. J.* 75, 484–492.
- (44) Hong, M., Fuangthong, M., Helmann, J. D., and Brennan, R. G. (2005) Structure of an OhrR-ohrA operator complex reveals the DNA binding mechanism of the MarR family. *Mol. Cell* 20, 131–141.
- (45) Kessler, A., Sezonov, G., Guijarro, J. I., Desnoues, N., Rose, T., Delepierre, M., Bell, S. D., and Prangishvili, D. (2006) A novel archaeal regulatory protein, Stal, activates transcription from viral promoters. *Nucleic Acids Res.* 34, 4837–4845.
- (46) Kim, J. Y., Inaoka, T., Hirooka, K., Matsuoka, H., Murata, M., Ohki, R., Adachi, Y., Fujita, Y., and Ochi, K. (2009) Identification and characterization of a novel multidrug resistance operon, mdtRP (yusOP), of *Bacillus subtilis*. *J. Bacteriol.* 191, 3273–3281.

Characterization of ligand and DNA binding abilities of BldR2 and BldR

Binding of BldR2 to selected target genes

In most archaea, other than the proposed autoregulatory mechanism of most MarR proteins, their physiological role remains elusive. To examine this point, have been investigated two issues: whether BldR2, like its homologue BldR, targets the MarR-like operon (*Sso1351-Sso1352*) and the *Sso2536* gene (involved in detoxification from aromatic compounds), and whether BldR2 is able to bind to sequences in the promoters of its neighbouring genes. The genes located nearby and upstream in the genome are the divergent *Sso1078* and *Sso1080* (which encode two components of an ABC-type antimicrobial peptide transport system), and a nearby downstream gene, *Sso1083* (which encodes the subunit A of the glutaconate CoA-transferase) (Fig. 11).

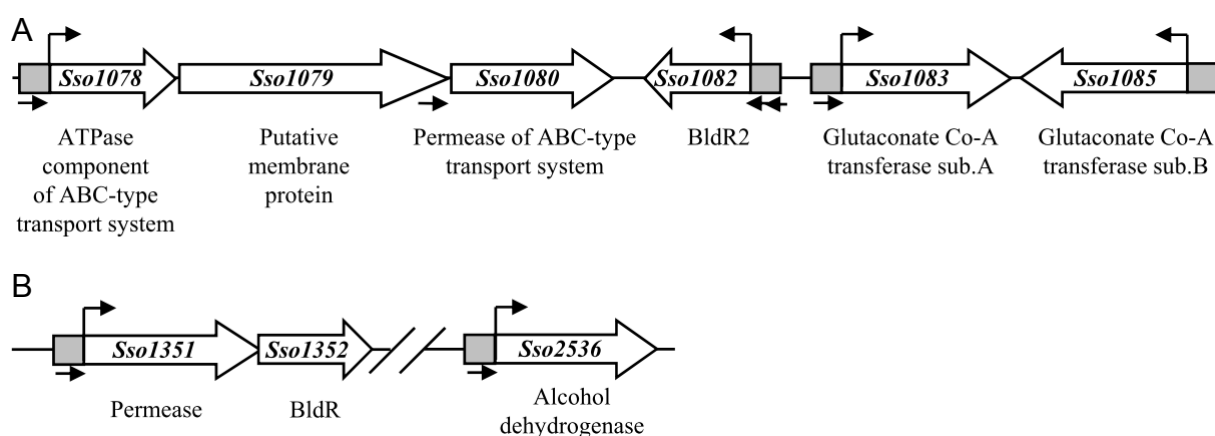


Fig. 11. Schematic representation of the genomic environment of *Sso1082*. ORFs are represented by arrowed boxes. From left to right: *Sso1078*, *Sso1079*, *Sso1080*, *Sso1082*, *Sso1083* and *Sso1085* are depicted with indications of their putative promoter regions (shaded rectangles). Thick bent arrows indicate the transcription start sites.

Hence, to identify putative BldR2 binding sites in the promoter regions of such genes, were amplified their 5' flanking regions by PCR, and the protein/DNA interaction was tested by EMSA. The three amplified sequences of *Sso1078*, *Sso1080* and *Sso1083*, were 162 bp, 156 bp and 115 bp long, respectively; they contained the predicted translation start sites, putative TATA/BRE sequences and the sequences immediately upstream, that usually contain regulatory binding sites. EMSA assays performed with purified BldR2 at saturating concentrations and DNA fragments containing the 5' flanking regions of *Sso2536* and *Sso1351* (of about 150 bp and 270 bp, respectively) revealed that BldR2 binds to both promoters (Fig. 12). Figure 12 also shows the ability of BldR2 to bind to sequences upstream of *Sso1078* and *Sso1080*, while the protein failed to recognize sequences upstream of *Sso1083*.

A recent report on the *S. solfataricus* transcriptome (reference 31 in the above-reported paper) has revealed that *Sso1078* and *Sso1080* are, together with *Sso1079*, encoding a putative S-layer domain protein, part of an operon. This means that the sequences upstream of *Sso1080*, recognized by BldR2 could not be promoter sequences even if a theoretical regulator binding site was found in this region. Furthermore, EMSAs revealed that BldR2 was able to bind to the promoter of *Sso2536* with an apparent K_d of 23.5 μ M (Fig. 13). This experiment was performed also with the two BldR2 mutants (R19A and A65S). Both mutants were able to

interact with target DNA and revealed affinity in the micromolar order (Fig. 13, Table 3). In particular, mutation R19A did not affect DNA binding ability, confirming that this amino acid is not critical for such ability. Instead, the A65S mutant showed a 20 fold increased affinity, indicating that the introduction of the serine 65 determined a strengthened DNA binding.

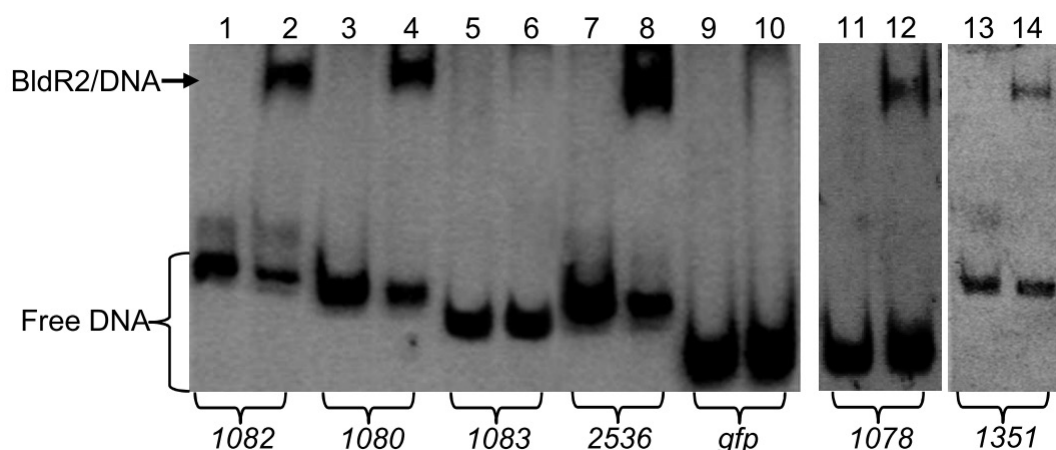


Fig. 12. Binding of BldR2 to the putative promoter regions of *Sso1082*, *Sso1080*, *Sso1083*, *Sso1078*, *Sso2536* and *Sso1351*. EMSA results using DNA fragments: 1082 (lane 1), 1080 (lane 3), 1083 (lane 5), 2536 (lane 7), *GFPfrag* (lane 9), 1078 (lane 11) and 1351 (lane 13) incubated with 62.5 μ M of purified BldR2 (lanes 2, 4, 6, 8, 10, 12 and 14, respectively). *gfp* represents the negative control.

The introduction of this serine acted in a way that the dissociation constant calculated towards two of its target promoters (*bldR2* and *adh* promoters) was more similar to that calculated for BldR (1.0 μ M versus its own promoter and 0.8 μ M versus *Sso2536*) than that of BldR2. Since a single substitution is able to influence protein function so dramatically, it could be concluded that this approach would be very helpful to correlate differences between the primary structure of BldR and BldR2 to their different *in vivo* physiological roles.

Taken together, these results indicate that besides being autoregulatory, BldR2 targets the *Sso1078*, *Sso1351* and *Sso2536* genes, suggesting that it likely regulates transcription of such stress response genes.

Table 3. K_d values of BldR2, R19A and A65S

	BldR2	R19A	A65S
<i>Sso1082prom</i>	28 μ M	16.2 μ M	1.2 μ M
<i>Sso2536prom</i>	23.5 μ M	9 μ M	0.7 μ M

Ligand binding of BldR2 and BldR

The majority of the characterized MarR-type regulators bind phenolic compounds to regulate either their metabolism or their export from the cell via efflux pumps.

To evaluate BldR2 and BldR conformational changes associated with benzaldehyde (BDH) and salicylate binding and to investigate their ligand specificity, the intrinsic fluorescence, CD spectroscopy and isothermal titration calorimetry (ITC) were used. The intrinsic fluorescence of BldR2 as a function of benzaldehyde concentration was analyzed to determine protein binding to this ligand (Fig. 14). Since BldR2 contains only one Trp residue, was used an excitation wavelength of 280 nm and ligand concentrations from 0.3 a 400 μ M. The fluorescence quenching (Q_{obs}/Q_{max}) at 320

nm plotted against ligand concentration generated a one-site binding curve with an apparent dissociation constant of 88 μM (Fig. 14B).

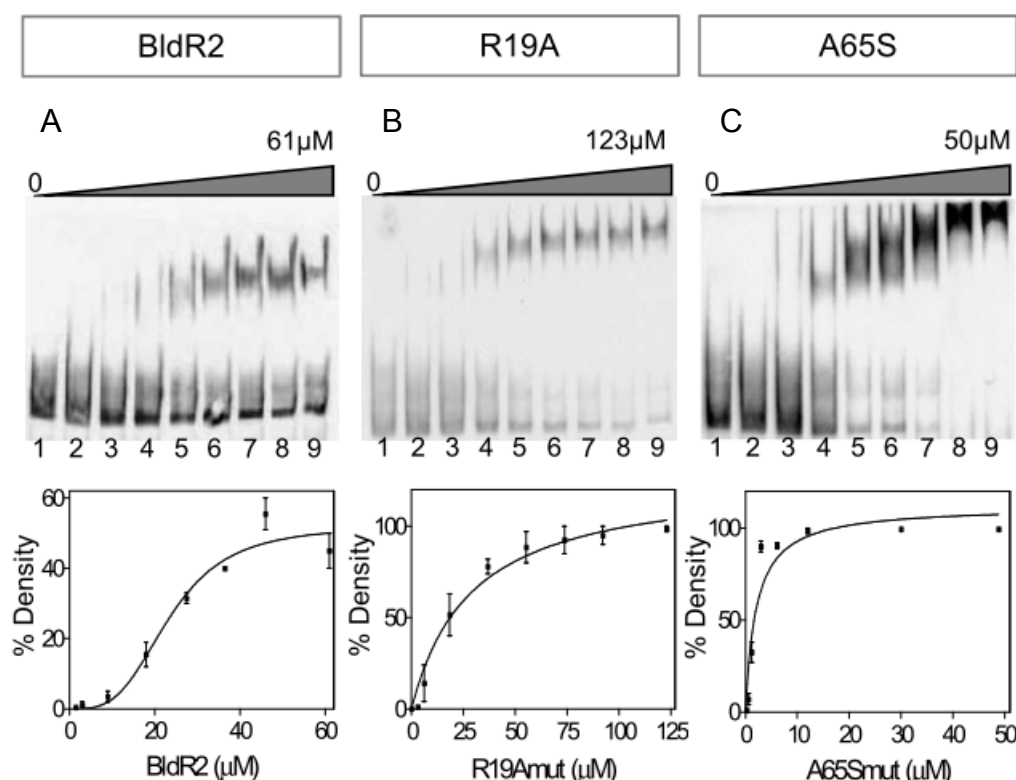


Fig. 13. Binding affinity of BldR2, R19A and A65S proteins to *adh* promoter region by EMSA. *Sso2536* promoter titrated with increasing concentrations of BldR2 (A), R19A (B) and A65S (C). Lane 1: DNA fragment; lanes 2–10, DNA probe incubated with protein from 1.5 μM to 61 μM for BldR2, from 1.5 μM to 123 μM for R19A and from 0.3 μM to 50 μM for A65S. The binding curves have been reported under the EMSAs; densitometric data from each EMSA are plotted against the concentrations of protein. Error bars represent the standard deviation for each point derived from triplicate experiments.

Unfortunately, the potential change in the protein conformation upon salicylate binding could not be monitored by spectrofluorimetry due to the strong emission by the ligand (at 360 nm). Fluorescence spectra were recorded in a Cary Eclipse spectrofluorimeter (Varian EN61010-1) at 25°C and 70°C using 9 μM BldR2 in 20 mM Tris-HCl pH 8.0. For each sample, three repetitive scans were obtained and averaged. Data elaboration was performed using the “Cary Eclipse” software, while saturation curves were obtained with the “GraphPad” software (Prism). The $Q_{\text{obs}}/Q_{\text{max}}$ values were calculated by $Q_{\text{obs}}/Q_{\text{max}} = (I_x - I_{\text{min}}) / (I_{\text{max}} - I_{\text{min}})$, where ‘I’ is the intensity of fluorescence.

Circular dichroism experiments showed that, in the presence of increasing concentrations of benzaldehyde, from 7.8 to 400 μM , the far-UV CD spectra of BldR2 increased in the negative ellipticity at 208 nm and 222 nm and decreased in the signal at 190 nm. These data indicate that significant conformational changes occur in the ligand-bound form at the secondary structural level (Fig. 15). The ellipticity at 222 nm was plotted versus the concentration of benzaldehyde and fitted with a one-site ligand-binding model, yielding a dissociation constant of 16 μM (Fig. 16).

Titration of BldR2 with salicylate also resulted in an effect on the negative ellipticities of the α -helices, clearly indicating a conformational change. The fitting of the binding

data with a one-site binding equation yielded a dissociation constant of 30 μM (Fig. 16).

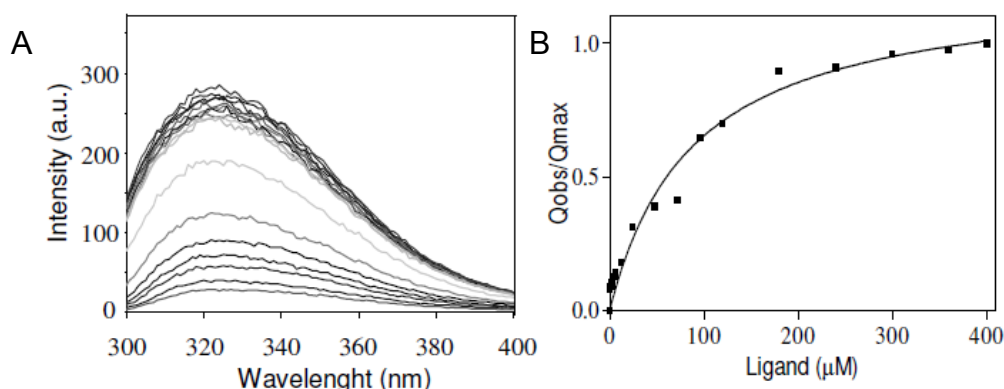


Fig. 14. Binding of BldR2 to benzaldehyde by fluorescence quenching analysis (A) Fluorescence spectra of BldR2 (9 μM) with varying concentrations of benzaldehyde. (B) Fluorescence quenching as a function of increasing concentrations of benzaldehyde. Data are the average of three independent experiments.

The same experiments were performed using chymotrypsinogen as a control protein and the benzaldehyde as the ligand. No variations in either molar ellipticity or fluorescence emission maximum were observed. The experiment was also performed with R19A and A65S proteins and the results are reported in Table 4.

CD spectra were obtained in a JASCO-J810 CD spectropolarimeter equipped with a temperature controller unit. FAR UV spectra were recorded from 260 to 190 nm using 31 μM BldR2 (and chymotrypsinogen) in 50 mM Na-phosphate pH 8.0. For each sample, three repetitive scans were performed and averaged. Depending on the conditions, a scan of the buffer or a scan of the buffer and ligand was subtracted from each of the protein spectra. Data elaboration was performed using the “Spectrum Analysis” software, while saturation curves were obtained with the “GraphPad” software (Prism).

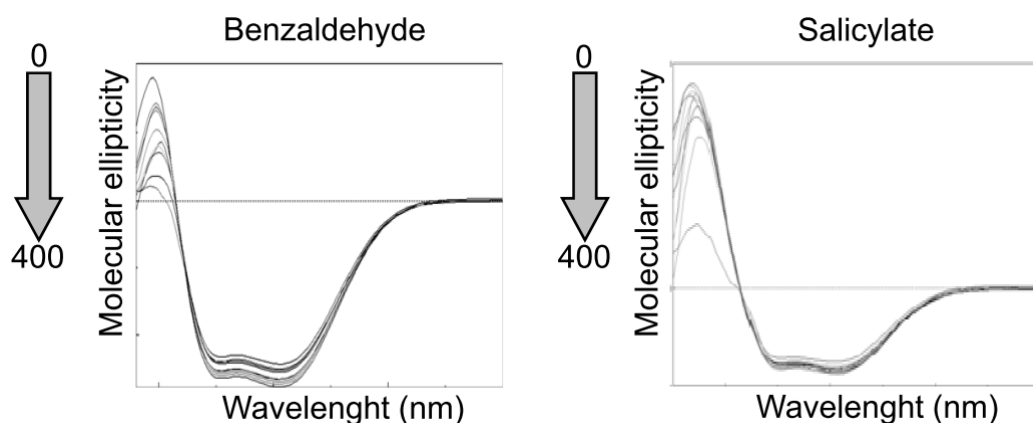


Fig. 15. Effect of benzaldehyde and salicylate on protein conformation by circular dichroism analysis.

From these results it has been proposed that the natural ligand for BldR2 could be benzaldehyde; in fact, using the intrinsic fluorescence analysis and the CD spectroscopy, it was proved that BldR2 can bind benzaldehyde with affinity constants

in the low micromolar range, which is consistent with a molecule having physiological relevance.

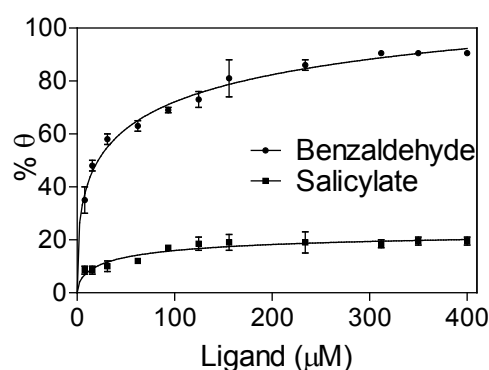


Fig. 16. Molar ellipticity percent as a function of increasing concentrations of benzaldehyde and salicylate; the percentages of molar ellipticity are calculated at 222 nm. Data are the average of three independent experiments.

As for BldR, its binding to the benzaldehyde was characterized by isothermal titration calorimetry (ITC). The titration of BldR, carried out in duplicate at 25° C, was performed using 12 μM of protein and a maximum ratio BldR/ligand equal to 1:10. The experiments were carried out using a N-ITC III *Calorimetry Sciences* instrument in 20 mM NaP pH 7.5, 0.5 M NaCl.

Table 4. Dissociation constants of BldR2, R19A and A65S versus benzaldehyde

	BldR2	R19A	A65S
BDH (CD)	16 μM	58 μM	48 μM
BDH (fluorescence)	88 μM	106 μM	98 μM

The titration with the ligand was carried out with a Hamilton syringe containing a stock solution of 480 μM of benzaldehyde, by 25 injections, each of 10 μL, at 200 seconds-time intervals. A control experiment was set up in the same conditions by performing a titration of the buffer alone. The ITC data were analyzed by using the NanoAnalyze program, from which the thermodynamic parameters were obtained. Figure 18 shows the raw heat rate in function of the time. The relationship between heat of association and number of moles of BDH added to each injection were plotted as a function of the molar ratio [BDH]/[BLDR] (Fig. 17). Using an independent binding site model available on the NanoAnalyze program the following thermodynamic parameters were obtained: K_a : $8.2(\pm 2.3) \times 10^5 \text{ M}^{-1}$; ΔH° : $31.9(\pm 1.0)$ kJ/mol; n : 6.

The value of the association constant between BldR and its allosteric modulator has a micromolar order, in agreement with the reversible association of the two partners. The positive ΔH° strongly suggests that, upon binding with the ligand, a strong conformational change occurs in the protein that requires energy. The stoichiometry (n) of the reaction indicates that there are three binding sites per monomer. The value of ΔG° , calculated with the equation $\Delta G^\circ = -RT \ln K_d$, is found to be -33.0 kJ/mol, demonstrating that the process is spontaneous. From the equation $\Delta G^\circ = \Delta H^\circ - T\Delta S^\circ$, a ΔS° of 0.22 kJ/mol was calculated. The data strongly suggest that the association between BldR and benzaldehyde is an endothermic process, that indicates that the binding results in a net change in the protein conformation. This is in agreement with the role of transcriptional activator played by BldR2 upon binding with the benzaldehyde.

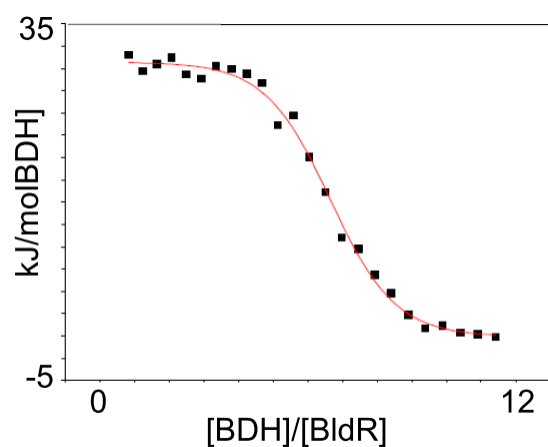
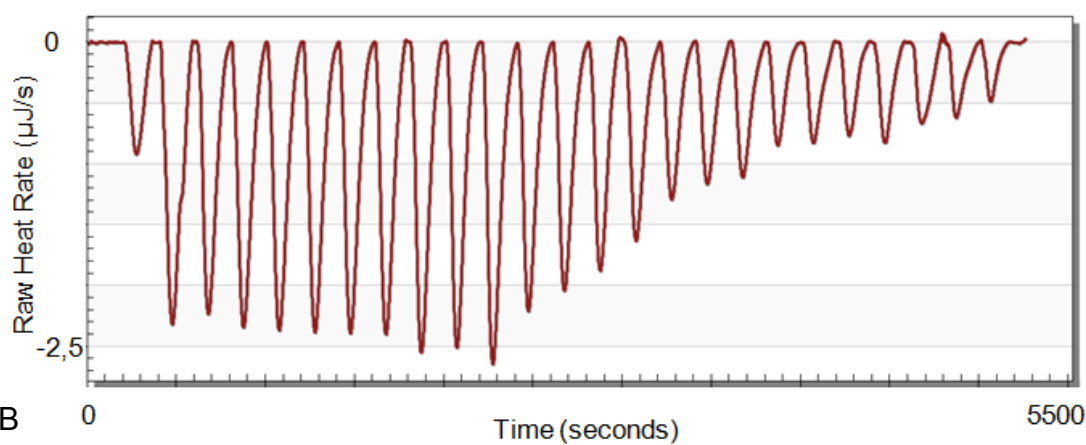


Fig. 17. The graph shows the heat of association/number of moles of BDH ratio, as a function of the [BDH]/[BldR] molar ratio. The number of benzaldehyde binding sites for each BldR dimer was derived from the graph (as the inflection point of the curve).

A



B

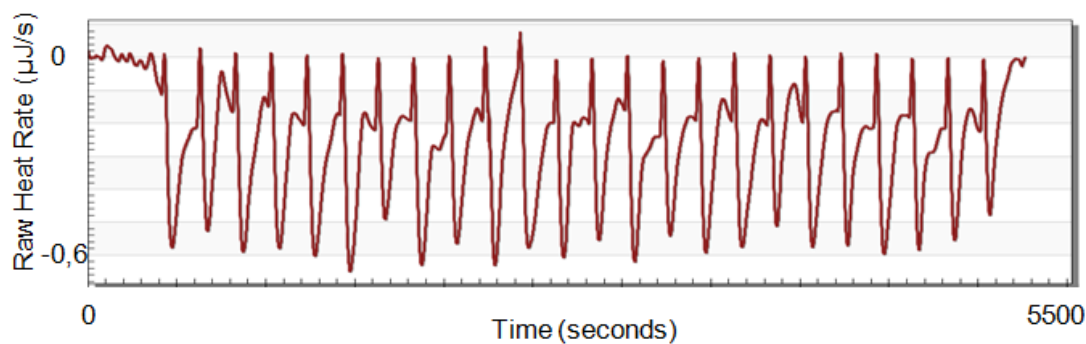


Fig. 18. (A) Isothermal titration calorimetric profile of BldR/benzaldehyde association. (B) Isothermal titration calorimetric profile of 20 mM NaP, NaCl 500 mM buffer with the benzaldehyde.

Chapter 3

STRUCTURAL AND FUNTIONAL CHARACTERIZATION OF TtArsC, A NOVEL ARSENATE REDUCTASE INVOLVED IN ARSENIC DETOXIFICATION IN THE BACTERIUM *THERMUS THERMOPHILUS* HB27

**A novel arsenate reductase from the bacterium *Thermus thermophilus* HB27:
its role in arsenic detoxification**

Immacolata Del Giudice^a, Danila Limauro^a, Emilia Pedone^{a,b}, Simonetta Bartolucci^a,
Gabriella Fiorentino^{a*}

^aDepartment of Biology, University of Naples Federico II, Edificio 7, via Cinthia,
80126 Naples, Italy

^bInstitute of Biostructures and Bioimaging, CNR, via Mezzocannone 16, 80134
Naples, Italy

*To whom correspondence should be addressed
fiogabri@unina.it;
phone +39081679167

Submitted to BBA Proteins and Proteomics

Abstract

Microorganisms living in arsenic-rich geothermal environments act on arsenic with different biochemical strategies, but the molecular mechanisms responsible for the resistance to the harmful effects of the toxic metalloid have only partially been examined. In this study, we investigated the mechanisms of arsenic resistance in the thermophilic bacterium *Thermus thermophilus* HB27. This strain, originally isolated from a Japanese hot spring, exhibited a good tolerance to high concentrations of arsenate and arsenite; it owns in its genome a putative chromosomal arsenate reductase (*TtarsC*) gene encoding a protein homologous to the well characterized one from the plasmid pI258 of the Gram⁺ bacterium *Staphylococcus aureus*. Differently from the majority of microorganisms, *TtarsC* is part of an operon including genes not related to arsenic resistance; qRT-PCR showed that its expression was four-fold increased when arsenate was added to the growth medium. Gene cloning and expression in *Escherichia coli*, followed by purification of the recombinant protein, proved that, like *S. aureus*, TtArsC is indeed a thioredoxin-coupled arsenate reductase and also exhibited weak phosphatase activity. The catalytic role of the first cysteine (Cys7) was ascertained by site directed mutagenesis. These results identify TtArsC as the main actor in the arsenic resistance in *T. thermophilus* giving the first structural-functional characterization of a thermophilic arsenate reductase.

Highlights:

Occurrence of molecular mechanisms for facing the toxicity of arsenic in *T. thermophilus*.

Transcriptional regulation of a putative arsenate reductase gene by arsenic.

First investigation on a thermostable Trx-dependent arsenate reductase from a hyperthermophile.

Identification of Cys7 as a residue essential for the catalysis.

Keywords: Hyperthermophile - arsenate resistance - arsenate reductase - protein thermal stability

Abbreviations: *ars* (arsenic resistance system) - *aox* (arsenite oxidation system) - *arr* (arsenate respiratory reduction) - *TtarsC* (gene ID: 2776273) - TM (*Thermus* medium) - Trx (thioredoxin) - Tr (thioredoxin reductase) - pNPP (p-nitrophenyl phosphate) - MIC (minimal inhibitory concentration) - LMWP (low molecular weight phosphatases).

1. Introduction

Arsenic is a toxic metalloid that is widely distributed in the Earth's crust and is naturally present in the soil, water and air from trace levels up to hundreds of mg/kg (or mg/L). The geochemical cycling of arsenic is very complex, in fact, in addition to chemical and physical parameters, it involves biological factors. Microbial activities play critical roles in the geochemical cycling of arsenic because they can promote or inhibit its release from sediment materials [1], mainly by redox reactions [2]. The reduction of pentavalent arsenate [As(V)] to trivalent arsenite [As(III)] is the major reaction causing the release of arsenic from the mineral surfaces into groundwater; in fact, besides being more toxic, arsenite is the most mobile and common form of arsenic found in anaerobic contaminated aquifers [3]. The frequent abundance of arsenic in the environment has guided the evolution of multiple resistance strategies in almost all microorganisms. In fact, to counteract the deleterious effects of arsenic toxicity, they have evolved intriguing mechanisms which can include: reduced arsenic uptake by increased specificity of phosphate transporters [4,5]; transformation of the metalloid to less-toxic forms by methylation and adsorption [6,7,8]; arsenite oxidation (*aox* genes) [9]; respiratory arsenate reduction (*arr* genes); arsenic resistance by arsenite extrusion [arsenic resistance system (*ars*) genes]. Among *arr* genes, a periplasmic arsenate reductase is a component of the respiratory electron transport chain in which arsenate is the terminal electron acceptor [2,10]. *Ars* systems allow the cell to detoxify intracellular arsenate and are widely distributed among bacteria and archaea [11]. They have been well documented in several bacteria including *E. coli* [12], *S. aureus* [13] and *Bacillus subtilis* [14] and are composed of *ars* genes encoding proteins that catalyse the two-electron reduction of As(V), which enters the cell as a phosphate analogue, followed by As(III) removal from the cell by a proton-driven arsenite transporter [15]. *Ars* genes can be carried on plasmids or chromosomes and are often organized in operons; the two most common contain either five (*arsRDABC*) or three (*arsRBC*) genes. The *arsRDABC* operon is found on the plasmids of Gram⁻ bacteria, such as *E. coli* R773, whereas *arsRBC* operons are generally found on the plasmids of Gram⁺ bacteria such as *S. aureus* pl258, or on bacterial chromosomes [15]. The *arsR* gene encodes a trans-acting repressor of the ArsR/SmtB family involved in transcriptional regulation, *arsB* encodes an arsenite efflux pump that exports arsenite but not arsenate [16], and *arsC* encodes a cytoplasmic arsenate reductase that converts arsenate to arsenite. Where present, ArsA is an arsenite-stimulated ATPase, and ArsD, a metallochaperone that transfers trivalent metalloids to ArsA [17]. In addition to these well-studied components, a variety of *ars* clusters contains additional genes whose functions in arsenic resistance has not been clearly established [18]. For example, *arsH* gene has been found in almost all Gram⁻ bacteria that contain the *ars* operon, but its function is not known yet [19]. A distinct gene (*arsM*), involved in the arsenic methylation, has been recently identified in more than 120 different archaea and bacteria [19], and characterized in *Rhodopseudomonas palustris* [7]. While the basic strategy to cope with arsenic is almost conserved among microorganisms, the arsenate reductases involved in detoxification are not. To date, three families of arsenate reductases have been identified and characterized, which differ from each other in several of their physical and catalytic properties [20]. ArsC from *S. aureus* pl258 (EC 1.20.4.1) has three redox active cysteines (Cys10, Cys82, Cys89) critical for its enzymatic activity. Cys10 is responsible for the initial nucleophilic attack on arsenate, subsequently Cys82 attacks the covalent intermediate and forms an intramolecular disulfide bridge.

Cys89 completes the reaction by forming a Cys82-Cys89 disulfide bond, that can, in turn, be reduced by a sequence of coupled redox reactions involving thioredoxin reductase (Tr), thioredoxin (Trx) and NADPH [21]. The two independent arsenate reductases from *E. coli* R773 and Acr2p from *Saccharomyces cerevisiae* are different from *S. aureus* ArsC pI258 because arsenate reduction leads to an intermolecular Cys-glutathione disulfide that is then reduced by glutaredoxin [22]. The exploration of life in extreme environments and the analysis of As-contaminated sites has led to the isolation of a large number of hyperthermophiles able to reduce arsenate or oxidize the arsenite for their metabolism [23,24]. For example, in the *Thermus* genus, *T. aquaticus* and *T. thermophilus* have been found to rapidly oxidize arsenite [25] and the ability of *T. thermophilus* HR13 to use arsenate for respiration has been shown [26]. Starting from these studies, it has been hypothesized that arsenic resistance could have evolved in geothermal environments where microorganisms have been exposed to natural sources of arsenic. Inspection of the genomes of different thermophilic Bacteria and Archaea has revealed putative genes involved in arsenic detoxification. However, in most of these organisms, the arsenic resistance has only been characterized in terms of the ability to grow in the presence of high arsenic concentrations (over 250 mM of arsenate) [27]; to date, *ars* systems and their genetic determinants have been preliminarily characterized only in *Geobacillus kaustophilus* and *Acidithiobacillus caldus* [28,29]. Hence, a characterization of the molecular and biochemical mechanisms responsible for arsenic resistance in thermophilic microorganisms is still at an age of infancy. In this study, we determined the involvement of a putative *arsC* gene (*TtarsC*) in the arsenic resistance of the thermophilic bacterium *T. thermophilus* HB27, its induction by arsenate, and the structural-functional characterization of the recombinant arsenate reductase with the identification of the role of the first cysteine in the redox cascade. In addition this report provides the first characterization of the enzymatic activity and the conformational stability of a thermophilic arsenate reductase.

2. Materials and Methods

2.1 Bacterial strains and growth conditions

T. thermophilus HB27 strain was purchased from the DSMZ and was grown aerobically at 74°C in TM medium composed of 0.8% (w/v) tryptone, 0.4% (w/v) yeast extract, 0.2% (w/v) NaCl, 0.35 mM CaCl₂ and 0.4 mM MgCl₂ and buffered at pH 7.0. A frozen (-80°C) stock of *T. thermophilus* HB27 cells was streaked on a TM plate (solidified by the addition of 0.8 % Gelrite) and incubated at 74°C overnight. Single colonies that appeared on the plate were inoculated into TM liquid medium and shaken at 74°C overnight. In order to determine the tolerance limits of *T. thermophilus*, arsenate (in the form of arsenic acid potassium salt anhydrous) and arsenite (in the form of sodium meta-arsenite) were added to final concentrations ranging between 1.5-30 mM and 0.25-20 mM, respectively, after dilution of an exponentially growing culture up 0.08 OD_{600nm}. For RT-PCR analysis, RNA was prepared from *T. thermophilus* HB27 cells grown in the presence or absence of 12 mM and 8 mM of arsenate and arsenite, respectively. The cells were grown to ~0.5 OD_{600nm} corresponding to mid-logarithmic phase and harvested by centrifugation at 6000 rpm for 10 min (F34-6-38 rotor; Eppendorf). For qRT-PCR experiments, cultures of *T. thermophilus* HB27 were grown in 500 mL of TM medium; when the cell density reached 0.5 OD_{600nm}, aliquots (50 ml) were harvested 0, 15, 30, 45 and 60

min after the addition of NaAsO₂ and KH₂AsO₄ (Sigma) at final concentrations of 8 mM and 12 mM, respectively. At these times, aliquots of each culture were removed and immediately spun down, and pellets were kept at -80°C. *E. coli* strains were grown in Luria Bertani [30] medium at 37°C with a 100 µg/ml ampicillin, 50 µg/ml kanamycin or 33 µg/ml chloramphenicol as required.

2.2 DNA and RNA extraction

Genomic DNA was prepared following reported procedures [31]. Total RNA was extracted using a RNAeasy Mini Kit (Qiagen). The extracted RNA samples were then diluted to 0.2 mg/ml for DNase treatment with the Ambion® TURBO™ DNase according to the manufacturer's instructions.

2.3 End-point reverse transcription RT-PCR

RT-PCR reactions were carried out on 2 µg of DNaseI-treated RNAs using SuperScript III Reverse Transcriptase (Invitrogen) following the manufacturer's protocol. Specific oligonucleotides (*1500priextrv*, *1501intrv*, *1502priext*) were designed on the base of *T. thermophilus* HB27 gene sequences available at <http://www.ncbi.nlm.nih.gov/nucore/46197919> [32] using Primer3Plus (<http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi/>) and utilized as primers for the RT reactions. The reactions were incubated at 55°C for 1h to produce the first-strand cDNA, followed by incubation at 70°C for 15 min to denature the reverse transcriptase. A control without reverse transcriptase was included for each RNA sample to ensure that DNA contamination had no effect on mRNA detection. PCR reactions were performed using specific primer pairs (*1500priextrv* and *1499operfw*; *1501intrv* and *1500intfw*; *1502priext* and *1502dirprext*) by 35 amplification cycles at 94°C for 1 min, an annealing temperature specific for each primer set for 1 min, 72°C for 1 min, and a final extension at 72°C for 10 min. The products of each PCR were detected by agarose gel electrophoresis. The primer sequences are reported in Table 1.

2.4 Quantitation of *TtarsC* transcripts

RT(q)PCR was performed to quantitate *TtarsC* (gene ID: 2776273, *TTC1502*) transcripts in NaAsO₂ and KH₂AsO₄-treated *T. thermophilus* HB27 cells using 7300/7500 Real time PCR system (Applied Biosystems) and the Maxima SYBR Green qPCR Master Mix kit (Fermentas Life Sciences). Specific *TtarsC* cDNA was synthesized as described above using the primer *arsCrealrv* and amplified using *arsCrealfw* and *arsCrealrv* primers. The oligonucleotides were designed using Primer Express 2.0 software (ABI Biosystems) and amplified a 100-bp *TtarsC*-specific product. As an internal control in this experiment the 16S rRNA gene was used. Primers *16SThfw* and *16SThrv* amplified a 111-bp 16S rRNA gene product. For the amplification of *TtarsC*, 25 ng from the RT-reaction mixture were used whereas 5 ng were used to amplify 16S fragment. DNA contamination was tested including for each RNA sample a control without reverse transcriptase. Two independent experiments were performed, and each sample was tested in triplicate. PCR amplification followed a standard protocol performing 40 cycles at 95°C for 15 s, and an annealing temperature specific for each primer set for 30 s. The amplification data were analyzed using the Sequence Detection System software (Applied

Biosystems). Induction folds were calculated by the comparative Ct method [33]. The relative expression ratio of the target gene, *TtarsC*, vs. that of the 16S rRNA gene was calculated by the equation: $RQ=2^{-\Delta\Delta Ct}$ whereas $\Delta\Delta Ct = \Delta Ct_{\text{reference}} - \Delta Ct_{\text{target}}$ and $\Delta Ct = Ct_{\text{gene of interest}} - Ct_{\text{reference gene}}$.

2.5 Heterologous expression and purification of arsenate reductase

The gene encoding *TtarsC* from *T. thermophilus* HB27 was amplified by PCR from genomic DNA, using Taq DNA polymerase (Thermo Scientific) and the primers containing *NdeI* (*arsCfw*) and *HindIII* (*arsCrv*) sites to the 5' and 3' ends of the fragment, respectively. Amplified fragments were purified, digested with appropriate restriction enzymes, and cloned in the *NdeI/HindIII*-digested pET30c(+) vector. The sequence of the cloned fragment was shown to be identical to the original annotated sequence (<http://www.ncbi.nlm.nih.gov/gene/2776273>). Blast analysis was performed to establish similarities among the sequence of *TtarsC* (<http://www.uniprot.org/uniprot/Q72HI5>) and the sequences of the SwissProt Data Bank. The multiple-sequence alignment was obtained using CLUSTAL-W. *E. coli* BL21-CodonPlus(DE3)-RIL cells (Stratagene) transformed with pET30/*TtarsC* were grown in LB medium containing kanamycin (50 µg/ml) and chloramphenicol (33 µg/ml). When the culture reached 0.5 OD_{600nm}, protein expression was induced by the addition of 0.5 mM isopropyl-1-thio-β-D-galactopyranoside and the bacterial culture was grown for an additional 4 h. Cells were harvested by centrifugation, and pellets were lysed by sonication in 50 mM Tris-HCl [pH 8.5] buffer in an ultrasonic liquid processor (Heat system Ultrasonic Inc.). The lysate was centrifuged at 25000 g for 60 min (JA25.50 rotor; Beckman). The supernatant was heated to 80°C for 10 min, and denatured proteins were precipitated by centrifugation at 25000 g for 30 min at 4°C. The supernatant was loaded onto a Resource Q column (1 ml; GE healthcare) connected to an AKTA Explorer system (GE Healthcare) and equilibrated in 50 mM Tris-HCl (pH 8.5). The elution was carried out with 20 ml of a KCl gradient (0 to 0.8 M). Fractions were collected and analyzed by SDS-PAGE to detect the *TtarsC* protein. These fractions were pooled, concentrated by ultrafiltration using a YM10 membrane (Millipore), dialyzed against 50 mM Tris-HCl (pH 8.5), and loaded onto a Superdex 75 column (16 x 60 cm; GE Healthcare) equilibrated in 50 mM Tris-HCl (pH 8.5), 0.2 M KCl at a flow rate of 1 ml/min. Fractions containing the *TtarsC* protein were pooled, concentrated, dialyzed and stored at 4°C. To prevent the oxidation of *TtarsC*, 1 mM DTT was added to all buffers, which were nitrogen flushed for several minutes prior to use.

2.6 Construction, Expression and Purification of C7S mutant

Single-point mutation in the *TtarsC* gene was produced with the Quick Change Site-Directed Mutagenesis kit (Stratagene) that utilizes PfuUltra DNA polymerase, using pET30/*TtarsC* wild type plasmid as DNA template. To generate the C7S mutant, the forward primer (*arsCC7Sfw*) with its complementary reverse primer (underlined letter in the primer's sequence in Table 1 indicates the base pair mismatch) was used and the reactions were performed according to the manufacturer's instructions. The mutagenesis products were transformed into XL-1 Blue Cells. Single colonies were selected on LB plates containing kanamycin (50 µg/ml), and purified plasmid DNA was sequenced at Eurofins MWG Operon to confirm the insertion of the correct

mutation. The mutant C7S protein was expressed and purified by the same procedure already described for the wild-type TtArsC protein.

2.7 Analytical Methods for Protein Characterization

Protein concentrations were determined by the Bradford method [34]. Protein homogeneity was estimated by SDS-PAGE [12.5% (w/v) gels]. To determine the native molecular mass of the proteins, the purified proteins (1.5 mg/ml) were applied in a volume of 100 μ l to an analytical Superdex PC75 column (3.2 cm x 30 cm) connected to an AKTA Explorer system (GE Healthcare) equilibrated with 50 mM Tris HCl (pH 8.5), 0.2 M KCl, at a flow rate of 0.04 ml/min. The column was calibrated using a set of gel filtration markers (low range, GE Healthcare), including bovine serum albumin (67.0 kDa), ovalbumin (43.0 kDa), chymotrypsinogen A (25.0 kDa), and RNase A (13.7 kDa). For further structural information purified TtArsC and C7S were analyzed by gel filtration chromatography connected to MiniDAWN Treos light scattering system (Wyatt Technology) equipped with a QELS module (quasi-elastic light scattering) for mass value. 100 μ g sample (1 mg/ml) was loaded on a Wyatt WTC015-S5 column (7.8 x 30 cm), equilibrated in 20 mM Tris HCl (pH 8.0), 150 mM NaCl and 1 mM DTT. A constant flow rate of 0.5 ml/min was applied. Data were analyzed by using Astra 5.3.4.14 software (Wyatt Technology). The molecular masses of the wild type protein and C7S mutant were also determined using electrospray mass spectra recorded on a Bio-Q triple quadrupole instrument (Micromass, Thermo Finnigan, San Jose, CA) [35].

2.8 Expression and purification of Trx from *E. coli* and Tr from *S. solfataricus* (SsTr)

E. coli Trx and *S. solfataricus* SsTr were expressed and purified and their activity measured as already described [36,37].

2.9 Arsenate reductase activity assay

Trx and SsTr proteins were defrosted and dialyzed against 50 mM TrisHCl (pH 7.0). TtArsC or C7S mutant were dialyzed against 50 mM TrisHCl (pH 7.0) and 1 mM DTT. NADPH (Sigma) was dissolved in water to a stock concentration of 20 mM. Arsenate (KH_2AsO_4) was freshly dissolved in water at 50 mM. The assay buffer contained 50 mM potassium phosphate (pH 7.0), 2 mM EDTA, 0.05 mM FAD. The final assay mixture was prepared by diluting all components in the assay buffer to obtain 1 μ M TtArsC (or C7S mutant), 25 μ M Trx, 2 μ M SsTr and 0.25 mM NADPH, taking into account the subsequent addition of arsenate (250 μ M). The assays were carried out in a final reaction volume of 500 μ l at 60°C. Arsenate reduction coupled to NADPH oxidation ($A_{\epsilon_{340}} = 6220 \text{ M}^{-1}\text{cm}^{-1}$) was measured by following the decrease in absorption at 340 nm. Control reactions were performed as described above without SsTr protein or TtArsC protein, respectively.

2.10 Assay of phosphatase activity

Purified TtArsC (100-200 μ g) was incubated at 40°C with 10 to 20 mM p-nitrophenyl phosphate (pNPP; Sigma) in 20 mM Tris HCl pH 7.0. The dephosphorylation of pNPP ($\Delta\epsilon_{405} = 18,000 \text{ M}^{-1}\text{cm}^{-1}$) was measured by following the increase in absorption

at 405 nm. Each reaction was performed in a total volume of 500 μ l and was corrected for non-enzymatic pNPP dephosphorylation.

2.11 Thermal resistance

TtArsC thermoresistance was estimated at 80°C and 87°C. The purified enzyme (6 μ M) was incubated at the indicated temperatures for 15, 30, 45, 90 and 120 min in 50 mM potassium phosphate (pH 7.0) and then the residual arsenate reductase activity was measured as described above.

2.12 Circular Dichroism Measurements

CD spectra were recorded by using a Jasco J-815 CD spectrometer, equipped with a Peltier-type temperature control system (model PTC-423S/15). Cells with path lengths of 0.1 cm were used in the far-UV region. CD spectra were recorded with a time constant of 4 s, a 2 nm bandwidth, and a scan rate of 20 nm/min; the signal was averaged over at least three scans and baseline corrected by subtraction of a buffer spectrum. Spectra were analyzed for secondary structure using CD Deconvolution PRO and Dichroweb softwares [38]. CD measurements were carried out using protein concentrations of about 12 μ M in a 25 mM Na-P (pH 7.0) buffer. The thermal unfolding curves were recorded in the temperature mode, by following the change in the CD signal at 222 nm with a scan rate of 1.0°C/min (range of temperature from 30°C to 105°C). A commercial 8 M GuHCl solution from Sigma was used as a denaturant solvent. TtArsC and C7S protein samples for GuHCl-induced denaturation experiments were incubated at 4°C for 1 day with the following increasing amounts of denaturing agent: 0.5 M, 1.5 M, 2 M and 3 M GuHCl. For each concentration two independent experiments were performed. The reversibility of the reaction was then checked by lowering the temperature. For each protein a value for T_m in the absence of the denaturant agent was estimated via linear extrapolation.

Table 1

Primer's name	Primer's sequence
1499operfw (A)	5'-GAGGCCATCAACAGTAGGG-3'
1500priextrv (B)	5'-TTCCCAAAGGCGGGGTCC-3'
1500intfw (C)	5'-CTTCGTGGGCCTCGCCGAC-3'
1501intrv (D)	5'-GGTGGCCTCCTCCAGGAAG-3'
1502dirprext (E)	5'-CCCTCCGCTTCGGCTTTGA-3'
1502priext (F)	5'-TTGGCCTCCTCCTTACGAA-3'
arsCrealfw	5'-GGAAACCCCTGGAGGAGTG-3'
arsCrearlv	5'-TCGTCGCTGGGAAGCCTTC-3'
16SThfw	5'-TAGTCCACGCCCTAAACGAT-3'
16SThrv	5'-CCTTTGAGTTTCAGCCTTGC-3'
arsCfw	5'-GGCTAGGCTTTGGCATATGCGGGTCCTCGTCC-3'
arsCrv	5'-GCGGCCATCCACGAAGCTTACAGGCCCTAAAGC-3'
arsCC7Sfw	5'-GTCCTCGTCCTCTCCACCCACAACCTCC-3'

3. Results

3.1 Growth of *T. thermophilus* HB27 in the presence of arsenic

As a first step to identify the genetic basis responsible for arsenic resistance in *T. thermophilus* HB27, we determined the MICs of NaAsO_2 and KH_2AsO_4 ; they were 15 mM and 20 mM respectively, indicating a high level of resistance to arsenic compounds (Fig S1). Analysis of the genomic sequence revealed an ORF, *TTC1502*, encoding a putative arsenate reductase (*TtArsC*). The genomic environment of *TTC1502* is represented in Fig. 1A and includes a putative gene cluster from *TTC1499* to *TTC1502*; upstream and downstream, ORF *TTC1494* and *TTC1503* are inversely oriented. *TTC1499* and *TTC1500* encode hypothetical proteins of 444 and 81 amino acids, respectively and are separated by 99 bp; *TTC1501* encodes a putative oligoendopeptidase F of 563 amino acids and the intergenic region upstream is of 25 bp; *TTC1502* is separated by *TTC1501* by 25 bp. Based on the sequence analysis it seemed that the genes responsible for arsenic resistance were not clustered in a “resistance” operon as the majority described in the literature [13]. From the NCBI database other homologues of known *arsC* genes were not found in *T. thermophilus*.

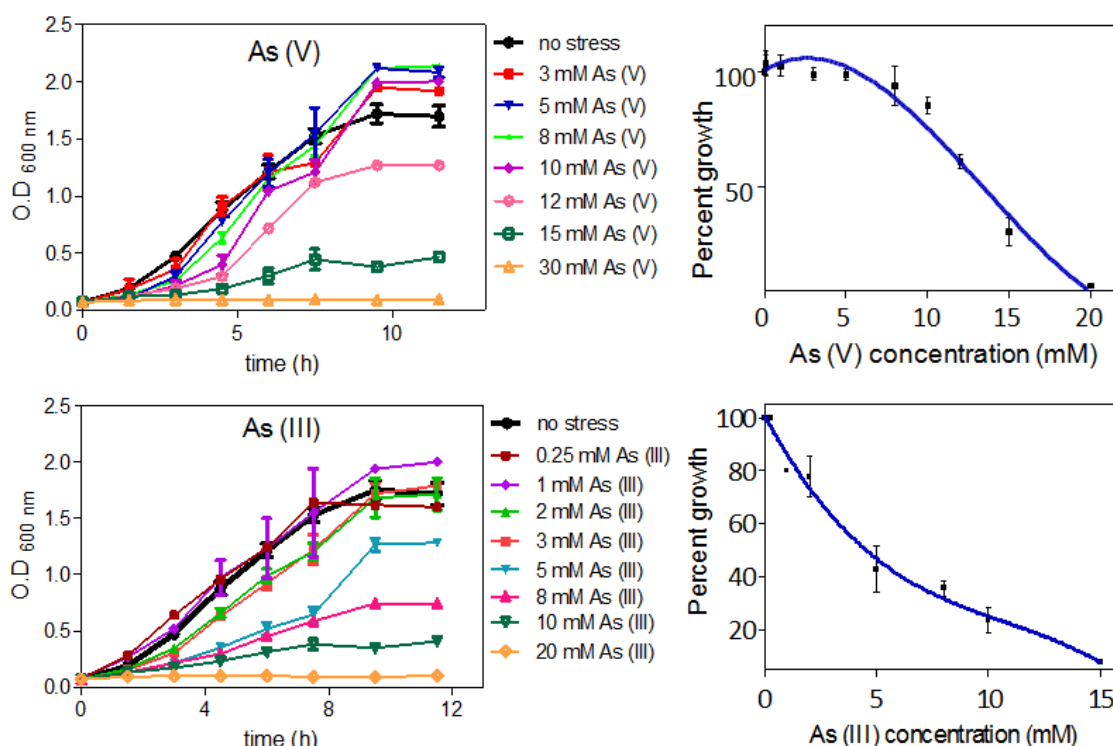


Fig. S1. Growth curves of *T. thermophilus* in the presence of increasing concentrations of arsenate and arsenite. Legend shows the arsenite and arsenate concentrations used for the calculation of the MICs. MIC values were calculated by plotting the percentage of growth against the concentrations of the metalloid.

3.2 Transcription of *TtarsC* in *T. thermophilus* HB27 and regulation by arsenic

To verify the transcription of the putative *TtarsC* gene, a RT-PCR was carried out using as templates DNaseI treated total RNAs extracted from cells grown up to 0.5

OD_{600nm} and exposed to 8 mM NaAsO₂ and 12 mM KH₂AsO₄ for 0, 15, 30, 45, 60 min and primers (*arsCfw*, *arsCrv*) annealing to the *TtarsC* coding sequence from position +1 to +453 with respect to the putative translation initiation site. *TtarsC* was expressed under all the conditions tested (data not shown).

In order to examine whether a single transcript was formed with contiguous upstream genes, as suggested by bioinformatics, we performed a RT-PCR analysis using for each reaction primer sets spanning the junctions between two adjacent genes (Table 1). In particular, with the above mentioned RNAs, a reverse transcription reaction was performed using alternatively the primer *1502priext* (F in Fig. 1A) annealing to a sequence in the *TtArsC* gene at +146 from the putative start codon; the primer *1501intrv* (D in Fig. 1A) annealing to a sequence in the *TTC1501* gene at +174; the primer *1500priextrv* (B in the Fig. 1A) annealing from position +165 in the corresponding gene. The three cDNAs obtained were then used in PCR reactions with primers *1499operfw* (A) and *1500priextrv* (B) which would amplify a fragment of 381 if the two genes *TTC1499* and *TTC1500* were cotranscribed. Failure to obtain PCR products from each of the cDNA indicated independent transcription of the two genes (Fig. 1B lanes 4, 6, 10). On the contrary, primers *1501intrv* (D) and *1500intfw* (C) in Fig. 1A amplified a 368 bp fragment when cDNAs D or F were used (Fig. 1B lanes 8, 12) showing that the two corresponding genes, *TTC1500* and *TTC1501* were cotranscribed; primers *1502priext* (F) and *1501dirpriext* (E) in Fig. 1A also amplified a 275 bp fragment indicating co-transcription of *TTC1501* and *TTC1502* (Fig. 1B lane 14). Taken together the results suggest that *TTC1500*, *TTC1501* and *TTC1502* are transcribed as a polycistronic messenger initiated from a promoter putatively located in the 99 bp intergenic region between *TTC1499* and *TTC1500*.

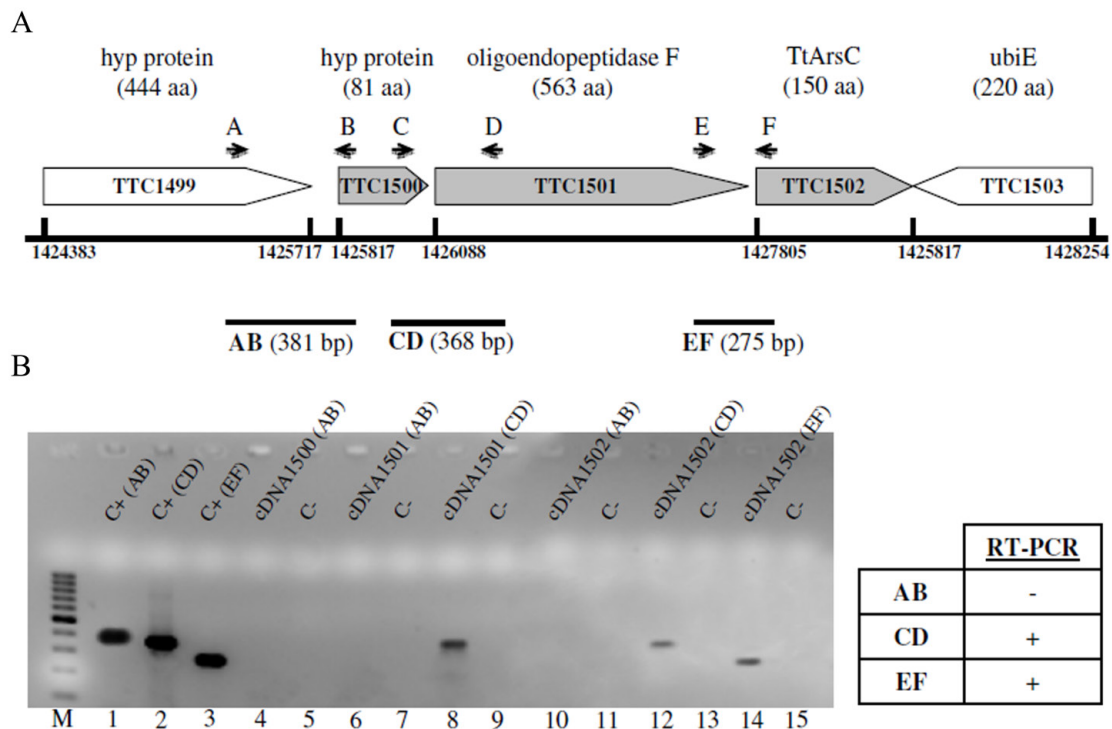


Fig. 1. (A) Schematic representation of the genomic environment of *TtarsC*. A diagram of the genomic region includes the *TTC1499*, *TTC1500*, *TTC1501*, *TtarsC* and *TTC1503* genes. Arrowed boxes depict each ORF and the direction of transcription, with the locus tag given inside each box. Arrows above boxed ORFs depict annealing positions and orientation of the primers used. The sequences of each primer are shown in Table 1. (B) Identification of the transcriptional unit. Agarose gel of RT-PCR

products. All RT-PCR products had the expected size (AB: 381 bp, CD: 368 bp, EF:275 bp). M: ladder 1Kb plus; lanes 1, 2, 3: PCR products amplified from *T. thermophilus* genome; lanes 5, 7, 9, 11, 13, 15: negative controls obtained using as template digested RNAs incubated without reverse transcriptase, lane 4: AB fragment from the 1500 cDNA; lane 6: AB fragment from the 1501 cDNA; lane 8: CD fragment from the 1501 cDNA; lane 10: AB fragment from the 1502 cDNA; lane 12: CD fragment from the 1502 cDNA; lane 14: EF fragment from the 1502 cDNA.

Investigation on the occurrence of similar gene association in annotated bacterial genomes through Blast analysis showed an identical organization only in *T. thermophilus* HB27 and HB8, whereas association of the putative oligoendopeptidase and the arsenate reductase was found not only in diverse *T. thermophilus* strains (JL18, SG0.5JP17-16, CCB_US3_UF1) but also in *T. oshimai*. To verify if gene expression was arsenic dependent, quantitative transcription of *TtarsC* was carried out in cells treated with arsenic compounds in comparison with untreated cells. As shown in Fig. 2, slight induction was observed in cells exposed to arsenite As(III), whereas a strong induction was measured in arsenate As(V) treated cells which reached a four-fold increase after 60 min of exposure to the arsenate. These results evidence a stringent response to arsenate, the substrate of the arsenate reductase, and strongly suggest an involvement of *TtarsC* in arsenate detoxification.

3.3 Cloning, expression and purification of *TtArsC*

To demonstrate that *TtArsC* was indeed an arsenate reductase we cloned the gene in pET30c(+) plasmid and expressed it in *E. coli* BL21-CodonPlus(DE3)-RIL. The translated sequence encodes a protein of 150 amino acids (16957.49 Da; pI 6.53) which contains three cysteine residues, Cys7, Cys83 and Cys90 conserved among the bacterial thioredoxin-coupled arsenate reductases which play a key role in the catalytic mechanism of the enzyme.

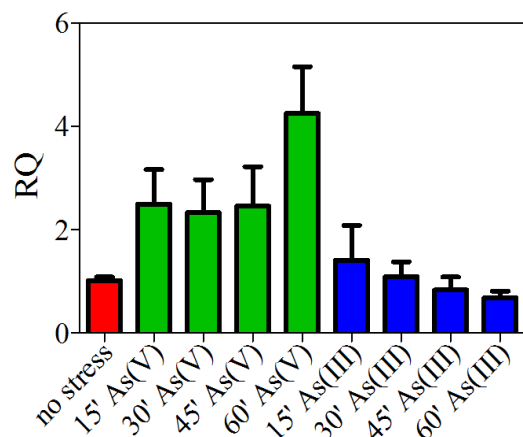


Fig. 2. Expression of *TtarsC* determined by qRT-PCR after stress by arsenate As(V) and arsenite As(III) at sub-inhibitory concentrations (12 mM and 8 mM) for 0, 15, 30, 45 and 60 minutes. Gene expression was normalized to that of the 16S rRNA gene. Error-bars represent SD of two independent experiments each performed in triplicate.

Sequence analysis and alignments with homologues also showed the presence of the anion binding site known as P-loop and the Asp-Pro sequence in which Asp serves as a catalytically important acid-base in LMWP (Fig. 3). The high identity with the annotated PTPs and the conservation of the residues involved in the phosphatase activity strongly suggest a dual phosphatase and reductase activity for *TtArsC*. Enzymes belonging to this family use the first Cys (included in the CX₅R motif) to form a thioester bond with the substrate arsenate as an early intermediate, while the two other cysteine residues are involved in a redox cascade; at first a disulfide bond between Cys7 and Cys83 is formed to cause the release of arsenite,

subsequently a disulfide bond between Cys83 and Cys90 is restored. The recycling of the enzymes is guaranteed by the intracellular Tr/Trx reducing system. Recombinant protein was purified to homogeneity through three steps: a heat treatment of the cell extract followed by anion exchange and gel filtration chromatographies (Fig. 4A). From 1 liter culture about 5 mg of pure protein were obtained. The expected molecular weight was confirmed by mass spectrometry.

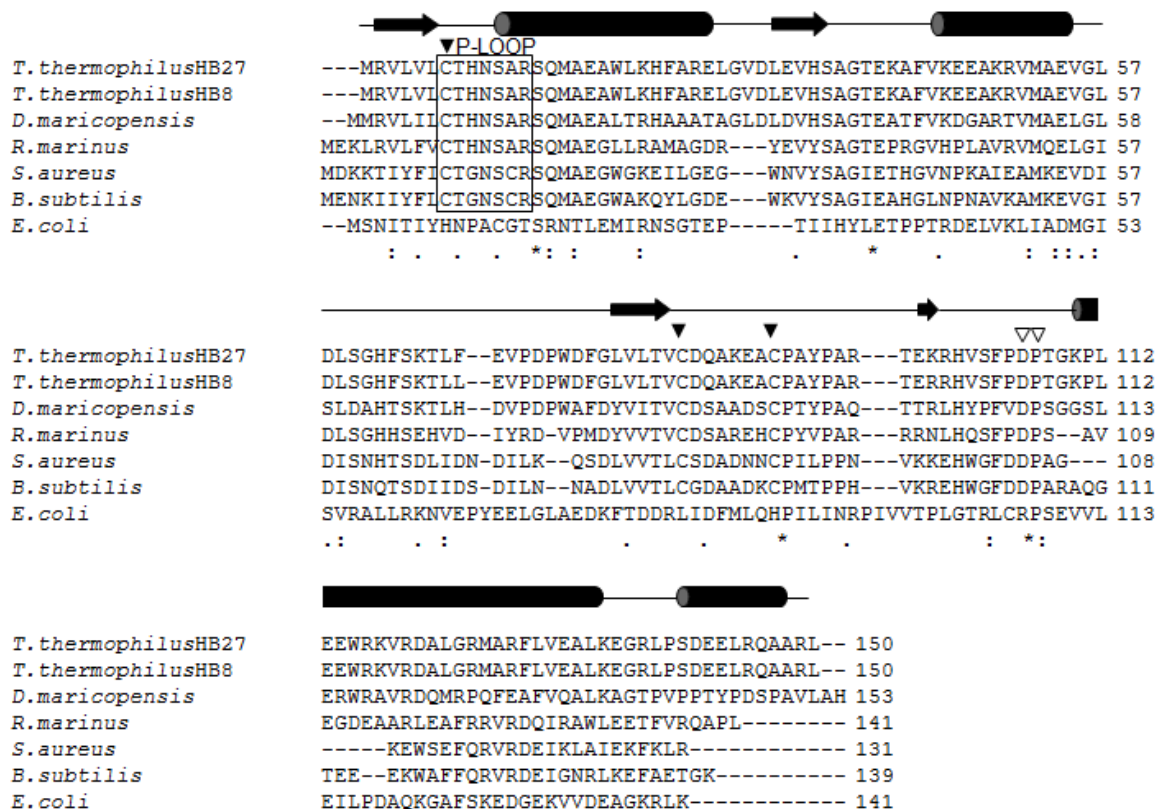


Fig. 3. Multiple sequence alignment by Clustal W of TtArsC (TTC1502) with bacterial arsenate reductases and bacterial PTPase homologues. Proteins are: ArsC (TTHA1853) of *T. thermophilus* HB8, PTPase (Deima_0085) of *Deinococcus maricopensis*, PTPase (Rmar_0375) of *Rhodothermus marinus*, ArsC (SAP018) of *S. aureus*, ArsC (BSSC8_16700) of *B. subtilis* and ArsC (b3503) of *E. coli* R773. The CX₅R motif is highlighted by a box. Filled arrowheads refer to the catalytic Cys7, Cys83 and Cys90. Asp-Pro sequence is indicated by open arrowheads. The secondary structure elements of TtArsC are depicted above the sequences.

In order to assess the structure of the recombinant protein, gel filtration chromatography coupled with light scattering methodology was also performed showing that the protein in solution is a monomer (Fig. 4B).

3.4 Functional characterization of TTC1502

There isn't a suitable method to measure arsenate reductase activity by monitoring the arsenite formation. The enzymatic activity can be measured indirectly by coupling the oxido/reducing system NADPH/Tr/Trx to TtArsC and following the decrease in NADPH at 340 nm [39]. To deliver reducing equivalents required to recycle TtArsC, we used an hybrid (and heterologous) system represented by the recombinant thermophilic SsTr of *S. solfataricus* and the recombinant Trx from *E. coli* which were purified as described [36,37]. In such system TtArsC was able to link the reduction of

arsenate to the consumption of NADPH (Fig. 5), clearly suggesting an arsenate reduction mechanism employing the Tr/Trx system for the redox recycling.

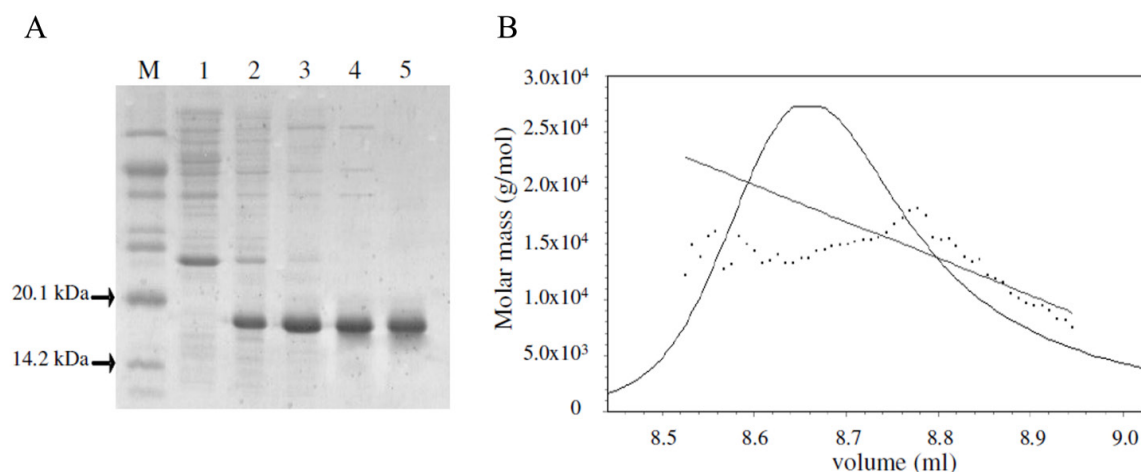


Fig. 4. (A) Analysis of recombinant TtArsC. SDS-PAGE of the purification steps: lane M, molecular mass marker; lane 1, cell extract from non-induced cells; lane 2, crude extract; lane 3, heat-treated cell extract; lane 4, fraction from anion exchange chromatography; lane 5, fraction from exclusion molecular chromatography. (B) Analysis of purified TtArsC by gel filtration chromatography coupled with in-line three-angle light scattering. The elution profile of ArsC is shown as a continuous line. The clustered points represent light scattering data converted to molecular mass.

At the temperature of 60°C, and using an arsenate saturating concentration of 0.25 mM, TtArsC enzyme had a rate of 4 $\mu\text{mol}/\text{min}$ per mg of protein, corresponding to about 68 turnovers per minute. To further analyze the reduction reaction mechanism, we investigated the contribution of the Cys7 in TtArsC catalytic activity by constructing a mutant in which the cysteine was substituted by a serine residue. The mutated gene was overexpressed in *E. coli* BL21-CodonPlus(DE3)-RIL and the corresponding protein was purified using the procedure described for the wild-type enzyme.

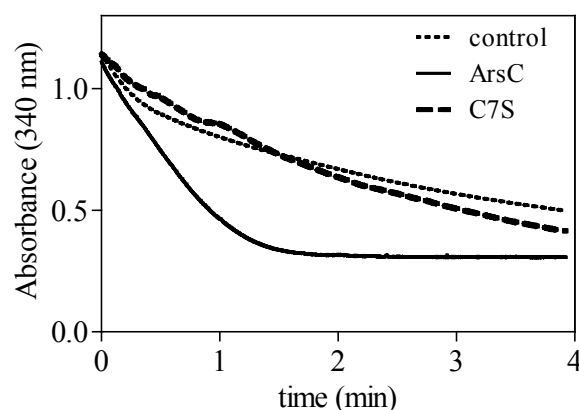


Fig. 5. (A) Arsenate reductase activity of TtArsC (—) and C7S (---) evaluated as rate of NADPH oxidation. Control (·····) was considered as rate of NADPH oxidation in the absence of arsenate reductase.

To characterize molecular weight and quaternary structure, LC-MS and light scattering analyses were performed. The value of the molecular mass obtained was in perfect agreement with the corresponding theoretical values and with the mutation introduced. Furthermore, light scattering analysis showed that the mutant protein retains its monomeric structure. The enzymatic activity of the C7S mutant was assayed with the NADPH/SsTr/Trx reducing system and compared to that of the native enzyme. The undetectable activity of C7S proved that Cys7 is a catalytic residue, presumably performing the nucleophilic attack on the arsenate, as already

described for Cys10 of *S. aureus* pI258 ArsC [20] (Fig. 5). We also tested the enzyme for phosphatase activity. Using a pNPP concentration of 10-20 mM, TtArsC exhibited a low, but detectable phosphohydrolase activity with a rate of 4 nmol/min per mg.

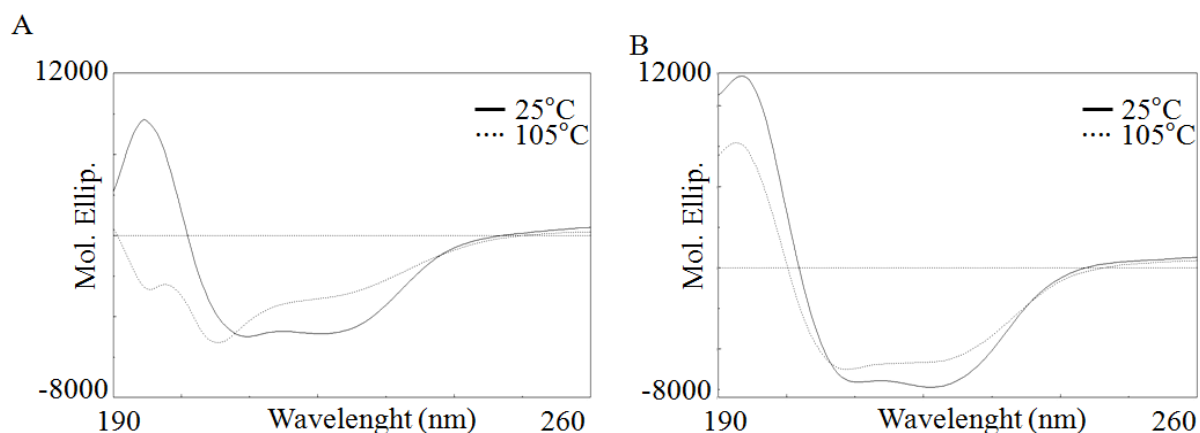


Fig. 6. Far-UV CD spectra of TtArsC (A) and C7S (B). Spectra were recorded at 25°C (—) and 105°C (····). The protein concentration used to acquire the spectra was 12 μ M.

3.5 Structural- functional characterization of TtArsC and C7S

To assess the secondary structure of TtArsC and its mutant and to compare the stabilities of the mutant and wild-type proteins at different temperatures, far-UV CD spectra in the 190–260 nm region were recorded in 25 mM Na-P buffer (pH 7.0) at 25°C and 105°C and the magnitude of the CD band at 222 nm was monitored at different temperatures. As shown in Fig. 6, the CD spectra of native TtArsC (A) and the mutant (B) exhibited two negative peaks at 208 and 222 nm and one positive peak at 195 nm, indicative of a predominantly folded structure with a α - β content. The α -helical and beta content of the wild-type protein was estimated using Dichroweb 2.0 and corresponded to 28% and 20%, respectively. The CD spectra of the mutant C7S indicated that the single mutation did not affect the overall secondary structure.

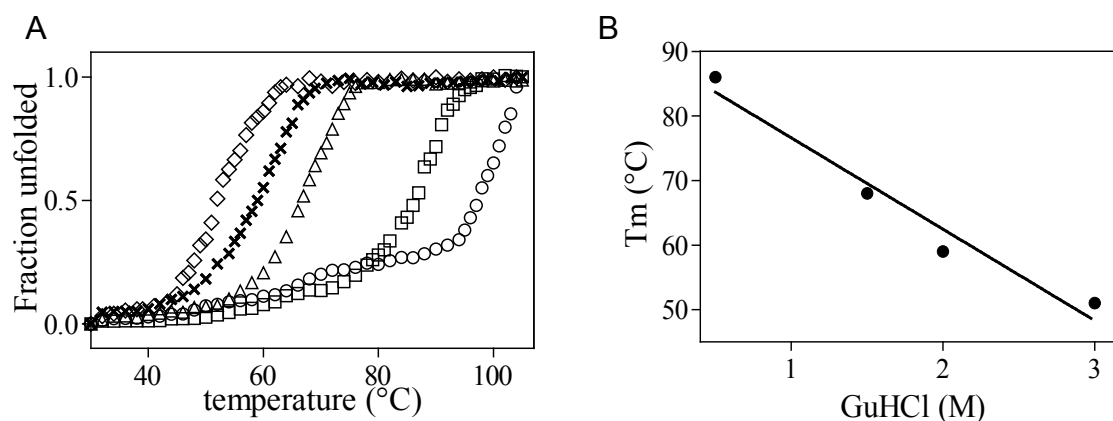


Fig. 7. (A) Thermal denaturation curves of TtArsC at a fixed concentration of 12 μ M, in 25 mM Na-P (pH 7.0) in the presence of 0 (O), 0,5 (\square), 1,5 (Δ), 2 (x) and 3 (\diamond) M GuHCl. The curves were obtained by recording the changes in the molar ellipticity at 222 nm as a function of temperature. (B) Dependence of the melting temperature T_m on GuHCl concentration; the T_m value in the absence of the denaturing agent was estimated via linear extrapolation.

Both proteins are endowed with high heat stability, in particular the C7S mutant CD spectrum recorded at 105°C displayed a greater preservation of secondary structure than the wild type protein. Measurements of residual activity of TtArsC after heating also proved high thermal stability; in fact the enzyme was fully active after 90 min at 80°C, and still maintained 50% activity after 30 min at 87°C (Fig S2).

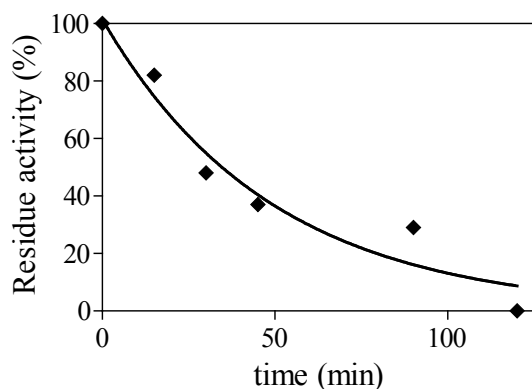


Fig. S2. TtArsC thermoresistance estimated at 87°C. The purified enzyme (6 μ M) was incubated at the indicated temperature for 15, 30, 45, 90 and 120 min.

Changes in the far-UV CD signal at 222 nm were used to follow the thermal unfolding; because of the very high thermal stability, it was not possible to obtain a complete thermal unfolding curve as the transition is not yet completed at 105°C (that is the maximum operating temperature of the instrument). The process was shown to be reversible for both the proteins. To obtain complete thermal denaturation, solutions of TtArsC and C7S at a fixed final concentration of 12 μ M were incubated in the presence of DTT with increasing GuHCl concentrations to progressively destabilize the native states. The corresponding curve for the wild type protein is collected in Fig. 7, and the denaturation temperatures obtained were plotted via linear extrapolation up to 0 M GuHCl giving an estimated melting temperature of 91°C. Interestingly, in all the conditions tested the mutant resulted to be more stable than the wild-type; in fact the thermal denaturation temperature calculated for C7S (101°C) proved to be higher than that of the wild type.

Discussion

Increased environmental pollution by arsenic, and the recent discussed discovery of arsenate-based life forms [40] has directed many research activities towards an understanding of the molecular mechanisms involved both in metal resistance and in the phosphate-arsenate promiscuity of phosphate utilizing enzymes and transporters [41]. As thermophilic organisms are also part of peculiar geochemical cycles in hostile environments, they represent an interesting model system for defining the ability to cope with metal stress under rapidly changing conditions. In this study, we identified a chromosomal *arsC* gene that contributes to arsenic resistance in *T.thermophilus* HB27 and biochemically characterized the encoded arsenate reductase. *Thermus* species live in arsenic-rich geothermal environments and can both oxidize and reduce arsenic thus playing an important role in the speciation and bioavailability of arsenic in thermal environments [26,19]. Our culturing experiments demonstrated the ability of *T. thermophilus* to grow in the presence of NaAsO₂ and KH₂AsO₄ in a comparable range of concentrations to those of bacteria classified as highly arsenic resistant [42]. These data further supported evidence of the presence

of molecular mechanisms for facing the toxicity of arsenic. Inspection of *T. thermophilus* genome revealed the presence on the chromosome of an open reading frame, *TTC1502*, encoding for a putative arsenate reductase, homologous to the one well-studied from *S. aureus* plasmid pI258; the genome context was very different, since no obvious ORFs encoding for putative partner proteins involved in arsenic detoxification were found either in the vicinity of *TTC1502* or on *T. thermophilus* megaplasmid pTT27. Separated *ars* genes have been found not only in other *Thermus* species (eg. *oshimai*, *scotoductus*) but also in more phylogenetically distant bacteria [43,44]. Transcriptional analysis of *TtarsC* showed that the gene is expressed and is the third of an operon also encoding a putative protein of 81 amino acids and a putative oligoendopeptidase F of 563 amino acids. Compared to other bacteria, the operon is very atypical because it encodes two proteins with no sequence homology with the proteins usually encoded by characterized *ars* operons and no apparent metabolic relationship. The absence of a resistance operon that could contribute to the arsenic defense mechanism, together with the presence in the entire genome of a single putative arsenate reductase, makes investigation on its role even more intriguing. Quantification of *TtarsC* transcription indicated low expression levels that increased up to four fold after cell exposure to arsenate, demonstrating an involvement of the gene in the defense mechanism. The evidence that *TtarsC* expression and induction by arsenate were not so pronounced as one would expect considering the absence of other arsenate converting enzymes, could imply that the encoded enzyme would be highly efficient. To investigate this point, TtArsC was overexpressed in *E. coli*, purified to homogeneity and structurally-functionally characterized. Recombinant TtArsC is a monomer of 150 amino acids with the highly conserved active-site sequence motif and the LMW PTPase domain characteristic of ArsC from *S. aureus* pI258 [22] and *B. subtilis* [45]; the secondary and the predicted tertiary structures are also shared with those members. We proved that TtArsC enzyme is able to convert arsenate to arsenite using electrons coming from the Tr-Trx system, with a catalytic mechanism involving the thiol group of the N-terminal Cys residue (Cys7), which performs a nucleophilic attack on the arsenate. The assignment of Cys7 as the first nucleophile was confirmed by the inability of C7S mutant to complete the Trx-coupled reaction cycle. Accordingly to the catalytic mechanism, *T. thermophilus* genome analysis showed the prevalence of putative Trxs and Trs (six and three, respectively) in comparison to glutaredoxin-like proteins (one), suggesting that in this organism the main source of reducing equivalents is represented by the Tr/Trx system. Furthermore, TtArsC also retained a weak phosphohydrolase activity, indicating its evolution from LMW PTPases and confirming the functional correlation with ArsC from *S. aureus* pI258. Intriguingly, TtArsC is the only characterized arsenate reductase from a Gram⁻ bacterium with such a dual catalytic activity (for example *E. coli* R773 ArsC has no measurable phosphatase activity) and it is one of the few described from Gram⁻ bacteria with a reduction mechanism characteristic of Gram⁺ [46]. TtArsC enzyme had an apparent V_{max} of 4 μmol/min per mg of protein, corresponding to about 68 turnovers per minute. The *in vitro* assay conditions, indicative of a vigorous enzyme, cannot even be considered optimal because heterologous electron donors, like SsTr and the Trx from *E. coli*, were used and the assay at a temperature of 60°C was performed, more likely below the predictable optimum; therefore, it is expected that *in vivo* TtArsC activity may be even higher than that of 6,4 and 19 turnovers per minute reported for both *S. aureus* pI258 and *E. coli* R773, respectively [47,48]. Another unique feature of TtArsC is its thermostability. Its high T_m (91°C) and long activity half-life at high

temperatures (higher than 90 min at 80°C) make it the most thermostable arsenate reductase described to date. In fact, *T. thermophilus* HB27 is the first thermophilic bacterium in which an ArsC activity has been characterized. Considering that arsenate reductases are relatively small proteins and that the identity with ArsC from *S. aureus* pl258, whose 3D structure is known, is of 40%, some determinants responsible for the higher thermostability of *TtArsC* could be identified, as, for example, an increase in proline residues content [49,50]. Interestingly, we found a higher thermal stability in the C7S mutant, but, as no structural information is yet available, it is difficult to draw conclusions about the structural and physical basis for this increased thermostability. In *S. aureus* the catalytic cysteine is located in a loop, presumably being extremely exposed to the solvent; hence, it can be speculated that the stabilization might occur upon removal of a reactive/destabilizing residue. Taken together, our data reveal that the physiological function of *TtArsC* could be a significant contribution to the arsenic resistance. Once entered into the cell, arsenate would increase *TtarsC* expression, probably through a still unidentified arsenic-dependent transcriptional regulator. As the arsenate is reduced by the enzyme, the even more toxic arsenite produced could be extruded by three putative arsenite permeases, two encoded by genes located on the chromosome and one encoded by a gene carried on the plasmid which have not been functionally characterized yet. Even though *TtarsC* expression levels are quite low, microbial cells could take benefit from both an efficient encoded enzyme and a finely regulated export outside the cell. The contribution of the other proteins to arsenic tolerance remains to be determined and will be investigated in a near future. Comprehensive knowledge of the molecular and genetic basis of arsenic metabolism and detoxification, besides being stimulating from an evolutionary point of view, represents an important starting point for developing efficient and selective arsenic bioremediation approaches, an environmentally friendly way for heavy-metal removal.

Acknowledgements

We acknowledge Dr. Raffaele Ronca for kind discussion and useful advice, Dr. Angelina Cordone for helpful assistance in the qRT-PCR experiments. Dr. Altea Alfè is also acknowledged for her helpful technical assistance.

Reference

- [1] C. Reyes, R. Lloyd, C.W. Saltikov, Geomicrobiology of Iron and Arsenic in Anoxic Sediments, in: Satinder Ahuja, Arsenic contamination of groundwater: Mechanism, Analysis, and Remediation, John Wiley & Sons Inc., New Jersey, 2008, pp. 123-146.
- [2] R.S. Oremland, J.F. Stolz, Arsenic, microbes and contaminated aquifers, Trends Microbiol. 13 (2005) 45-49.
- [3] J.N. Murphy, C.W. Saltikov, The ArsR Repressor mediates arsenite-dependent regulation of arsenate respiration and detoxification operons of *Shewanella* sp. Strain ANA-3, Journal of Bacteriology 191 (2009) 6722-6731.
- [4] C. Cervantes, G. Ji, J.L. Ramirez, S. Silver, Resistance to arsenic compounds in microorganisms, FEMS Microbiol. Rev. 15 (1994) 355-367.
- [5] C. Murota, H. Matsumoto, S. Fujiwara, Y. Hiruta, S. Miyashita, M. Shimoya, I. Kobayashi, M.O. Hudock, R.K. Togasaki Norihiro, N. Sato, Arsenic tolerance in a *Chlamydomonas* photosynthetic mutant is due to reduced arsenic uptake even in light conditions, Planta 236 (2012) 1395-1403.
- [6] R. Turpein, M. Panssar-Kallio, M. Haggblom, T. Kairesalo, Influence of microbes on the mobilization, toxicity and biomethylation of arsenic in soil, Sci. Tot. Environ. 236 (1999) 173-180.

- [7] J. Qin, B.P. Rosen, Y. Zhang, G. Wang, S. Franke, C. Rensing, Arsenic detoxification and evolution of trimethylarsine gas by a microbial arsenite S-adenosylmethionine methyltransferase, *Proc. Natl. Acad. Sci. U.S.A.* 103 (2006) 2075-2080.
- [8] R. Say, N. Yilmaz, A. Denizli, Biosorption of cadmium, lead, mercury, and arsenic ions by the fungus *Penicillium purpurogenum*, *Sep. Sci. Technol.* 38 (2003) 2039-2053.
- [9] V. Andreoni, R. Zanchi, L. Cavalca, A. Corsini, C. Romagnoli, E. Canzi, Arsenite oxidation in *Ancylobacter dichloromethanicus* As3-1b strain: detection of genes involved in arsenite oxidation and CO₂ fixation, *Curr. Microbiol.* 65 (2012) 212-218. 14
- [10] I. Rauschenbach, E. Bini, M.M. Häggblom, N. Yee, Physiological response of *Desulfurispirillum indicum* S5 to arsenate and nitrate as terminal electron acceptors, *FEMS Microbiol. Ecol.* 81 (2012) 156-162.
- [11] S. Bartolucci, P. Contursi, G. Fiorentino, D. Limauro, E. Pedone, Responding to toxic compounds: a genomic and functional overview of Archaea, *Frontiers in Bioscience* 18 (2013) 165-189.
- [12] C.W. Saltikov, B.H. Olson, Homology of *Escherichia coli* R773 *arsA*, *arsB*, and *arsC* genes in arsenic-resistant bacteria isolated from raw sewage and arsenic-enriched creek waters, *Appl. Environ. Microbiol.* 68 (2002) 280-288.
- [13] G. Ji, S.J. Silver, Regulation and expression of the arsenic resistance operon from *Staphylococcus aureus* plasmid pI258, *Bacteriol.* 174 (1992) 3684-3694.
- [14] T. Sato, Y. Kobayashi, The *ars* operon in the skin element of *Bacillus subtilis* confers resistance to arsenate and arsenite, *J. Bacteriol.* 180 (1998) 1655-1661.
- [15] P.C. Patel, F. Goulhen, C. Boothman, A.G. Gault, J.M. Charnock, K. Kalia, J.R. Lloyd, Arsenate detoxification in a *Pseudomonad* hypertolerant to arsenic, *Arch. Microbiol.* 187 (2007) 171-183.
- [16] D. Gatti, B. Mitra, B.P. Rosen, *Escherichia coli* soft metal ion-translocating ATPases, *J. Biol. Chem.* 275 (2000) 34009-34012.
- [17] Y.F. Lin, A.R. Walmsley, B.P. Rosen, An arsenic metallochaperone for an arsenic detoxification pump, *PNAS* 103 (2006) 15617-1562.
- [18] A. Achour-Rokbani, A. Cordi, P. Poupin, P. Bauda, P. Billard, Characterization of the *ars* gene cluster from extremely arsenic-resistant *Microbacterium* sp. strain A33, *Appl. Environ. Microbiol.* 76 (2010) 948-955.
- [19] D. Pérez-Espino, J. Tamames, V. de Lorenzo, D. Cánovas, Microbial responses to environmental arsenic, *Biomaterials* 22 (2009) 117-130.
- [20] J. Messens, S. Silver, Arsenate reduction: thiol cascade chemistry with convergent evolution, *J. Mol. Biol.* 362 (2006) 1-17.
- [21] Y. Li, Y. Hu, X. Zhang, H. Xu, E. Lescop, B. Xia, C. Jin, Conformational fluctuations coupled to the thiol-disulfide transfer between thioredoxin and arsenate reductase in *Bacillus subtilis*, *Journal of biological chemistry* 282 (2007) 11078-11083.
- [22] I. Zegers, J.C. Martins, R. Willem, L. Wyns, J. Messens, Arsenate reductase from *S. aureus* plasmid pI258 is a phosphatase drafted for redox duty, *Nature Structural Biology* 8 (2001) 843-847.
- [23] J.S. Blum, T.R. Kulp, S. Han, B. Lanoil, C.W. Saltikov, J.F. Stolz, L.G. Miller, R.S. Oremland, *Desulfohalophilus alkaliarsenatis* gen. nov., sp. nov., an extremely halophilic sulfate- and arsenate-respiring bacterium from Searles Lake, California, *Extremophiles* 16 (2012) 727-742.
- [24] A. Volant, C. Desoeuvre, B. Casiot, S. Lauga, G. Delpoux, J.C. Morin, M. Personné, F. Héry, P.N. Elbaz-Poulichet, O. Bertin, Bruneel, Archaeal diversity: temporal variation in the arsenic-rich creek sediments of Carnoulès Mine, France, *Extremophiles* 16 (2012) 645-657.
- [25] T.M. Gihring, G.K. Druschel, R.B. McCleskey, R.J. Hamers, J.F. Banfield, Rapid arsenite oxidation by *Thermus aquaticus* and *Thermus thermophilus*: field and laboratory investigations, *Environ. Sci. Technol.* 35 (2001) 3857-3862.
- [26] T.M. Gihring, J.F. Banfield, Arsenite oxidation and arsenate respiration by a new *Thermus* isolate, *FEMS Microbiol. Lett.* 204 (2001) 335-340.
- [27] C.R. Jackson, K.G. Harrison, S.L. Dugas, Enumeration and characterization of culturable arsenate resistant bacteria in a large estuary, *Syst. Appl. Microbiol.* 28 (2005) 727-734.
- [28] M. Cuebas, D. Sannino, E. Bini, Isolation and characterization of arsenic resistant *Geobacillus kaustophilus* strain from geothermal soils, *J. Basic. Microbiol.* 51 (2011) 364-371.
- [29] M. Dopson, E.B. Lindström, K.B. Hallberg, Chromosomally encoded arsenical resistance of the moderately thermophilic acidophile *Acidithiobacillus caldus*, *Extremophiles* 5 (2001) 247-255.
- [30] J. Miller, Experiments in molecular genetics, Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., 1972.15
- [31] G. Fiorentino, R. Ronca, R. Cannio, M. Rossi, S. Bartolucci, MarR-like transcriptional regulator involved in detoxification of aromatic compounds in *Sulfolobus solfataricus*, *J. Bacteriol.* 189 (2007) 7351-60.

- [32] A. Henne, H. Brüggemann, C. Raasch, A. Wiezer, T. Hartsch, H. Liesegang, A. Johann, T. Lienard, O. Gohl, R. Martinez-Arias, C. Jacobi, V. Starkuviene, S. Schlenczeck, S. Dencker, R. Huber, H.P. Klenk, W. Kramer, R. Merkl, G. Gottschalk, H.J. Fritz, The genome sequence of the extreme thermophile *Thermus thermophilus*, *Nat. Biotechnol.* 22 (2004) 547-553.
- [33] M. Pfaffl, A new mathematical model for relative quantification in real-time RT-PCR, *Nucleic Acids Res.* 29 (2001) 2002-2007.
- [34] M.M. Bradford, A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding, *Anal. Biochem.* 72 (1976) 248-254.
- [35] D. Limauro, M. Saviano, I. Galdi, M. Rossi, S. Bartolucci, E. Pedone, *Sulfolobus solfataricus* protein disulphide oxidoreductase: insight into the roles of its redox sites. *Protein Eng. Des. Sel.* 22 (2009) 19-26.
- [36] E. Pedone, M. Saviano, M. Rossi, S. Bartolucci, A single point mutation (Glu85Arg) increases the stability of the thioredoxin from *Escherichia coli*, *Protein Engineering* 14 (2001) 255-260.
- [37] E. Pedone, D. Limauro, R. D'Alterio, M. Rossi, S. Bartolucci, Characterization of a multifunctional protein disulfide oxidoreductase from *Sulfolobus solfataricus*, *FEBS J.* 273 (2006) 5407-5420.
- [38] L. Whitmore, B.A. Wallace, *Biopolymers* 89 (2008) 392-400.
- [39] J. Messens, J.C. Martins, E. Brosens, K. Van Belle, D.M. Jacobs, R. Willem, L. Wyns, Kinetics and active site dynamics of *Staphylococcus aureus* arsenate reductase, *J. Biol. Inorg. Chem.* 7 (2002) 146-156.
- [40] F. Wolfe-Simon, J. Switzer Blum, T.R. Kulp, G.W. Gordon, S.E. Hoefft, J. Pett-Ridge, J.F. Stolz, S.M. Webb, P.K. Weber, P.C. Davies, A.D. Anbar, R.S. Oremland, A bacterium that can grow by using arsenic instead of phosphorus, *Science* 332 (2011) 1163-1166.
- [41] M. Elias, A. Wellner, K. Goldin-Azulay, E. Chabriere, J.A. Vorholt, T.J. Erb, D.S. Tawfik, The molecular basis of phosphate discrimination in arsenate-rich environments, *Nature* 491 (2012) 134-137.
- [42] L. Wang, B. Jeon, O. Sahin, Q. Zhang, Identification of an arsenic resistance and arsenic-sensing system in *Campylobacter jejuni*, *Appl. Environ. Microbiol.* 75 (2009) 5064-5073.
- [43] R. Li, J.D. Haile, P.J. Kennelly, An arsenate reductase from *Synechocystis* sp. strain PCC 6803 exhibits a novel combination of catalytic characteristics, *J. Bacteriol.* 185 (2003) 6780-6789.
- [44] X. Li, L.R. Krumholz, Regulation of arsenate resistance in *Desulfovibrio desulfuricans* G20 by an *arsRBCC* operon and an *arsC* gene, *J. Bacteriol.* 189 (2007) 3705-3711.
- [45] M.S. Bennett, Z. Guan, M. Laurberg, X.D. Su, *Bacillus subtilis* arsenate reductase is structurally and functionally similar to low molecular weight protein tyrosine phosphatases, *Proc. Natl. Acad. Sci. U.S.A.* 98 (2001) 13577-13582.
- [46] B.G. Butcher, S.M. Deane, D.E. Rawlings, The chromosomal arsenic resistance genes of *Thiobacillus ferrooxidans* have an unusual arrangement and confer increased arsenic and antimony resistance to *Escherichia coli*, *Appl. Environ. Microbiol.* 66 (2000) 1826-1833.
- [47] G. Ji, E.A. Garber, L.G. Armes, C.M. Chen, J.A. Fuchs, S. Silver, Arsenate reductase of *Staphylococcus aureus* plasmid pI258, *Biochemistry* 33 (1994) 7294-7299.
- [48] T.B. Gladysheva, K.L. Oden, B.P. Rosen, Properties of the arsenate reductase of plasmid R773, *Biochemistry* 33 (1994) 7288-7293.
- [49] G. Fiorentino, R. Cannio, M. Rossi, S. Bartolucci, Decreasing the stability and changing the substrate specificity of the *Bacillus stearothermophilus* alcohol dehydrogenase by single amino acid replacements. *Protein Eng.* 11 (1998) 925-930.
- [50] G. Fiorentino, I. Del Giudice, S. Bartolucci, L. Durante, L. Martino, P. Del Vecchio, Identification and physicochemical characterization of BldR2 from *Sulfolobus solfataricus*, a novel archaeal member of the MarR transcription factor family, *Biochemistry* 50 (2011) 6607-662.

Chapter 4

CHARACTERIZATION OF A PUTATIVE TRANSCRIPTIONAL REGULATOR, TtSmtB, INVOLVED IN ARSENIC DETOXIFICATION IN THE BACTERIUM *THERMUS THERMOPHILUS* HB27

4.1 Introduction

All cells require transition metal ions as cofactors in metalloenzymes or for structural or regulatory roles. All metal ions are toxic when they exceed their physiological levels, and hence their intracellular concentration is so tightly controlled that metal homeostasis allows organisms to rapidly respond to changes in their microenvironments [6]. Metal ions that don't play any biological role, such as heavy metal pollutants, must be detoxified via either biotransformation or efflux from the cytosol. The expression of heavy metal resistance genes is controlled at the transcriptional level by metal sensor proteins that 'sense' specific metal ions by forming specific coordination complexes [1] and modulate expression of genes that encode proteins that expel, sequester, or otherwise detoxify the excess metal. The MerR and SmtB/ArsR families represent two general classes of transcriptional regulators that have endowed prokaryotes with the ability to respond to stress induced by heavy metal-toxicity. The interest of this study is the SmtB/ArsR family, named after *E. coli* R773 ArsR [15], which regulates the *ars* operon expression in response to trivalent As(III) and Sb(III), and the Zn(II) sensor *Synechococcus* SmtB [7], which regulates expression of the *smtA* gene, encoding a class II metallothionein involved in sequestering excess Zn(II).

A distinguishing feature of ArsR/SmtB members is the presence of an ELCV(C/G)D motif, which is required for metal ion sensing; this motif, termed 'metal binding box', is highly conserved among the members of the family. The X-ray structure of apo-SmtB revealed that the ELCV(C/G)D motif is located in the $\alpha 3$ helix, as part of the projected $\alpha 3$ -turn- $\alpha 4$ DNA binding motif [4]. X-ray absorption spectroscopy of As(III)-ArsR complex revealed that As(III) is coordinated via three cysteine thiolates within the putative $\alpha 3$ helix, two of which derive from the ELCVCD motif [9]; substitutions of one or both cysteines with non-metal binding residues in the ELCVCD motif inhibits the ability of arsenic to dissociate ArsR from the *ars* O/P. Among SmtB/ArsR members there is a great structural diversity in the sensing site, in fact $\alpha 3N/\alpha 3$ [14], $\alpha 4C$ [3], $\alpha 5/\alpha 5C$ [14] and $\alpha 5-3$ [2] sites were identified.

SmtB/ArsR members bind to a DNA promoter containing one or two imperfect 12-2-12 inverted repeats, generally in proximity or overlapping the transcriptional start site. The *smtA* promoter, for example, contains two imperfect inverted 12-2-12 repeats (S2/S1 and S4/S3), the first of which is required for full Zn(II) responsiveness of *smtA* expression, whereas the second has little, if any, effect on the regulation of *smtA* [13]. The stoichiometry of O/P binding by SmtB/ArsR repressors remains elusive. It was supposed that a single homodimer would bind to a single inverted repeat, with the DNA recognition helices ($\alpha 4$) interacting with successive major grooves. The model which describes the mechanism of SmtB/ArsR repression is a derepression, in which metal binding by the sensor protein weakens the DNA binding affinity, so that RNA polymerase can load and initiate transcription of the operon.

Here we report the identification and characterization of a new sensor protein belonging to ArsR/SmtB family from *Thermus thermophilus* HB27. *T. thermophilus* is a valuable model organism for exploring arsenic resistance, because its genome contains a gene encoding an arsenate reductase (*TTC1502*) [5], two genes encoding putative arsenite efflux pumps (ArsB) (*TTC1447* and *TTP0033*) and four genes encoding homologues of ArsR/SmtB members (*TTC0353*, *TTC0617*, *TTC1830* and *TTC0296*). Among the putative transcriptional regulators found on the genome, *TTC0353* (*TtSmtB*) was chosen because it shares the highest identity with the known members of the ArsR/SmtB family. It has been demonstrated that *TtsmtB* gene is

transcribed in an operon of five genes which appear not to be metabolically correlated; a two-fold increase in *TtsmtB* expression was observed in cells exposed to arsenate and arsenite. The protein was expressed, purified and preliminary characterized. It was able to bind its own putative promoter and the putative promoter of *TtarsC* operon. When added to the samples, the arsenite was able to alleviate the binding of *TtSmtB* to the target promoter suggesting its role as a derepressor. Despite all attempts to increase the protein stability and the ability to manipulate it *in vitro*, it was not possible to perform additional characterizations to determine the quaternary structure and obtain a fine characterization of its DNA binding ability.

4.2 Materials and Methods

4.2.1 Plasmids, bacterial strains, and growth conditions

Strains and plasmids used in this study are described in Table 1 and 2.

Table 1

Strain	Genotype	Source
<i>Thermus thermophilus</i> HB27	Wild type	DSMZ (Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH)
<i>E. coli</i> TOP10F'	F' {lacI ^q Tn10 (Tet ^R)} mcrAΔ(mrr-hsdRMSmcrBC) φ80ΔlacZΔM15 ΔlacX74 recA1 deoR araD139 Δ(ara-leu) 7697galU galKtpsL(Str ^R) endA1 nupG	Invitrogen
<i>E. coli</i> BL21-Codon Plus (DE3) RIL	F– ompThsdS(r _B – m _B –) dcm+ Tet ^r gal λ (DE3) endAhte [argUileYleuWCam ^r]	Stratagene

Table 2

Plasmids	Source
pCR®4-TOPO	Invitrogen
pET28b(+)	Novagen
pET28b(+)- <i>TtsmtB</i>	This study
pET30c(+)	Novagen
pET30c(+)- <i>TtsmtB</i>	This study

T. thermophilus HB27 strain was grown as described in the previous chapter. For RT-PCR and qRT-PCR experiments, cultures of *T. thermophilus* HB27 were grown in 500 mL of TM medium; when the cell density reached 0.5 OD_{600nm}, aliquots (50 ml) were harvested at 0, 15, 30, 45 and 60 min after the addition of NaAsO₂ and KH₂AsO₄ (Sigma) at the final concentrations of 8 mM and 12 mM, respectively. At these times, aliquots of each culture were removed and immediately spun down, and pellets were kept at -80°C.

E. coli strains were grown as described in the previous chapter.

4.2.2 DNA and RNA extraction

Genomic DNA and total RNA were prepared as previously reported in chapter 3.

4.2.3 End-point reverse transcription RT-PCR

RT-PCR reactions were carried out on 2 µg of DNaseI-treated RNAs using SuperScript III Reverse Transcriptase (Invitrogen) following the manufacturer's protocol. Specific oligonucleotides (*0352rv*, *0353rv*, *0354rv* and *0355rv* reported in the Table 3) were designed based on *T. thermophilus* HB27 gene sequences using Primer3Plus software and used as primers for the RT reactions. The reactions were incubated at 55°C for 1h to produce the first-strand cDNA, followed by incubation at 70°C for 15 min to inactivate the reverse transcriptase. A control without reverse transcriptase was included for each RNA sample to ensure that DNA contamination had no effect on mRNA detection. PCR reactions were performed using specific primer pairs (*0351fw* and *0352rv*; *0352fw* and *0353rv*; *0353fw* and *0354rv*; *0354fw* and *0355rv*) by 35 amplification cycles of 94°C for 1 min, a specific annealing temperature for each primer set for 1 min, 72°C for 1 min, and a final extension at 72°C for 10 min. The products of each PCR were detected by agarose gel electrophoresis.

4.2.4 Quantitation of *TtsmtB* and *TTC0354* transcripts

To determine whether *TtsmtB* gene expression (*TTC0353*) was induced by arsenic, RT(q)PCR was performed to quantitate *TtsmtB* transcripts in NaAsO₂ and KH₂AsO₄-treated *T. thermophilus* HB27 cells using 7300/7500 Real time PCR system (Applied Biosystems) and the Maxima SYBR Green qPCR Master Mix kit (Fermentas Life Sciences). Total RNA was extracted using the RNeasy mini kit (Qiagen), according to the procedure supplied with the product, and treated with Turbo DNase, RNase free (Ambion), to eliminate DNA contamination in each sample. Two cDNAs were synthesized using a mixture of the proper reverse primer (*0353rv* or *0354realrv*) and the 16S reverse primer (Table 3), used as internal control. Specific *TtsmtB* cDNA was synthesized and amplified using *smtBrealfw* and *0353rv* or *0354realfw* and *0354realrv* primers or 16Sfw and 16Srv (Table 3). The oligonucleotides were designed using Primer Express 2.0 software (ABI Biosystems) and amplified a 107-bp *TtsmtB*-specific product or a 89-bp *TTC0354*-specific product. The 16S rRNA was used for the normalization as already described. For the amplification of *TtsmtB* and *TTC0354*, 25 ng from the RT-reaction mixture were used whereas 5 ng were used to amplify 16S fragment. DNA contamination was tested by the inclusion of a control without reverse transcriptase for each RNA sample. Two independent experiments were performed, and each sample was always tested in triplicate. PCR amplification followed a standard protocol, with a 15 s denaturation phase at 95°C, and a specific annealing temperature for each primer set for 30 s for 40 cycles. The amplification data were analyzed using the Sequence Detection System software (Applied Biosystems). Induction folds were calculated by the comparative Ct method. The relative expression ratio of the target gene, *TtsmtB* or *TTC0354*, vs. that of the 16S rRNA gene was calculated by the equation: $RQ = 2^{-\Delta\Delta C_t}$, whereas $\Delta\Delta C_t = \Delta C_t \text{ reference} - \Delta C_t \text{ target}$ and $\Delta C_t = C_t \text{ gene of interest} - C_t \text{ reference gene}$.

4.2.5 Cloning, heterologous expression and purification of TtSmtB

The gene encoding TtSmtB (*TTC0353*) from *T. thermophilus* HB27 was amplified by PCR on genomic DNA, using Taq DNA polymerase (Thermo Scientific) and the primers containing NdeI (*smtBfw*) and HindIII (*smtBrv*) sites at the 5' and 3' ends, respectively. Amplified fragments were purified and cloned in pCR®4-TOPO, according to the TOPO TA cloning (Invitrogen) kit instructions. pCR®4-TOPO/*TtsmtB*

was digested with appropriate restriction enzymes, and cloned in the NdeI/HindIII-digested pET30c(+) and pET28b(+) vectors. The strategy adopted produced recombinant plasmids able to produce the TtSmtB protein, fused (pET28) or not (pET30) to His-tag. The sequence of the cloned fragments was shown to be identical to the original annotated sequence (<http://www.ncbi.nlm.nih.gov/gene/2776273>) and, in the case of pET28/*TtsmtB*, confirmed the correct fusion of TtSmtB to the His₆ tag. Blast analysis was performed to establish similarities among the sequence of TtSmtB and the sequences of the SwissProt Data Bank. The multiple-sequence alignment was obtained using CLUSTAL-W. *E. coli* BL21-CodonPlus(DE3)-RIL cells transformed with pET28/*TtsmtB* and pET30/*TtsmtB* were grown in LB medium containing kanamycin (50 µg/ml) and chloramphenicol (33 µg/ml). For the pET28/*TtsmtB*-transformed cells, when the culture reached 0.7 OD_{600nm}, protein expression was induced by the addition of 0.5 mM isopropyl-1-thio-β-D-galactopyranoside (IPTG) and the bacterial culture was grown for an additional 4 h. Cells were harvested by centrifugation, and pellets were lysed by sonication in lysis buffer (50 mM Tris-HCl [pH 8.0], 1 mM phenylmethylsulfonyl fluoride, 1X inhibitor cocktail, 1 mM DTT) in an ultrasonic liquid processor (Heat system Ultrasonic Inc.). The lysate was centrifuged at 25000 g for 60 min (JA25.50 rotor; Beckman). The supernatant was heated to 70°C for 10 min, and denatured proteins were precipitated by centrifugation at 25000 g for 30 min at 4°C. The supernatant was loaded onto a HiTrap Heparin column (5 ml; GE Healthcare) connected to an AKTA Explorer system (GE Healthcare) and equilibrated in 50 mM Tris-HCl (pH 8.0). The elution was carried out with 60 ml of a KCl gradient (0 to 1 M). Fractions were collected and analyzed by SDS-PAGE to detect the TtSmtB protein. The fractions containing TtSmtB were pooled, concentrated by ultrafiltration, dialyzed using a HiTrap Desalting column (5 ml; GE Healthcare), equilibrated in 50 mM Tris-HCl (pH 8.0), 25 mM KCl at a flow rate of 5 ml/min. Fractions containing the TtSmtB protein were pooled, concentrated and stored at -80°C in single-use aliquots.

4.2.6 Analytical Methods for Protein Characterization

Protein concentrations were determined by the Bradford method (reference 34 in the chapter 3) and BCA protein assay kit (Pierce), with bovine serum albumin as a standard. Protein homogeneity was estimated by SDS-PAGE [12.5% (w/v) gels]. To determine the native molecular mass of the protein, the purified TtSmtB (1.5 mg/ml) was applied in a volume of 100 µl to an analytical Superdex PC75 column (3.2 cm x 30 cm) connected to an AKTA Explorer system (GE Healthcare) equilibrated with 50 mM Tris HCl (pH 8.0), 0.2 M KCl, at a flow rate of 0.04 ml/min. The column was calibrated using a set of gel filtration markers (low range, GE Healthcare), including bovine serum albumin (67.0 kDa), ovalbumin (43.0 kDa), chymotrypsinogen A (25.0 kDa), and RNase A (13.7 kDa). The molecular mass of the wild type protein was also determined using electrospray mass spectra recorded on a Bio-Q triple quadrupole instrument (Micromass, Thermo Finnigan, San Jose, CA) (reference 35 in the chapter 3). SDS-Page analysis of TtSmtB in the presence of 2-mercaptoethanol was performed; TtSmtB (8 µg) was incubated with 2 mM 2-mercaptoethanol for 10 minutes.

4.2.7 Circular Dichroism Measurements

CD spectra were recorded by using a Jasco J-815 CD spectrometer, equipped with a Peltier-type temperature control system (model PTC-423S/15). Cells with path lengths of 0.1 cm were used in the far-UV region. CD spectra were recorded with a time constant of 4 s, a 2 nm bandwidth, and a scan rate of 20 nm/min; the signal was averaged over at least three scans and baseline corrected by subtraction of a buffer spectrum. Spectra were analyzed for secondary structure using CD Deconvolution PRO and Dichroweb softwares (reference 38 in the chapter 3). CD measurements were carried out using protein concentration of about 15 μ M in a 25 mM Na-P (pH 8.0) buffer.

Table 3. Oligonucleotides

Primer name	Primer sequence
<i>smtBfw</i>	5' - GTCCAAGGAGGAGGAAACATATGCCAAGCGGGG - 3'
<i>smtBrv</i>	5' - GCATCATTTGAGCAAGCTTTCAAGTGTTCCTTCCGC - 3'
<i>0351fw</i>	5' - CACATGTCCCGCATCGCC - 3'
<i>0352rv</i>	5' - CCGCTCCAAGGCCATCAC - 3'
<i>0352fw</i>	5' - TGGACCAGCTCAAGGAGG - 3'
<i>0353rv</i>	5' - TCG CAGACGCAAAGCTCC - 3'
<i>0353fw</i>	5' - CCACCAGCTCAGGCTTCTC - 3'
<i>0354rv</i>	5' - CTCGCGTAGTTCACCTGGAC - 3'
<i>0354fw</i>	5' - GACCTCTGCTTTGGCTTTG - 3'
<i>0355rv</i>	5' - GCGGAAATCCCCCTGTAAG - 3'
<i>0354realfw</i>	5' - GACCTCTGCTTTGGCTTTG - 3'
<i>0354realrv</i>	5' - CTACCTGCCAACTCCTCCA - 3'
<i>16Sthfw</i>	5' - TAGTCCACGCCCTAAACGAT - 3'
<i>16Sthrv</i>	5' - CCTTTGAGTTTCAGCCTTGC - 3'
<i>0351promfw</i>	5' - GAAGCTTGAAGGGGGCTC - 3'
<i>0351promrv</i>	5' - GCGAGCCACAATACC - 3'
<i>1500promfw</i>	5' - TGAGGCGCAAGCTCATCCC - 3'
<i>1500promrv</i>	5' - TTCCCAAAGGCGGGGTCC - 3'
<i>1078fw</i>	5' - GGGCGCAGTATCGGTATATTTC - 3'
<i>1078rv</i>	5' - CATTCCCCTCATAGATCTTGGTAAC - 3'
<i>smtBrealfw</i>	5' - GAGAGGGTGGTCAAGGAG - 3'

4.2.8 DNA binding assays

To determine the binding of TtSmtB to the promoter regions of the *TTC0351/TTC0352/TTC0353/TTC0354/TTC0355* operon, containing the *TtsmtB* gene, and the *TTC1500/TTC1501/TTC1502* operon, containing the *TtarsC* gene, an electrophoretic mobility shift assay (EMSA) was performed. The 259-bp promoter region of the *TTC0351/TTC0352/TTC0353/TTC0354/TTC0355* operon and the 267-bp promoter region of the *TTC1500/1501/1502* operon were amplified using *0351promfw* and *0351promrv*, and *1500promfw* and *1500promrv* primers (Table 3). The purified DNA fragments were then labeled with [32 P]-dCTP by nick translation. The labeling reaction was performed in 25 μ l total volume at 15°C for 60 minutes as follows: 1mM dATP, dGTP, dTTP; 1X DNA polymerase Buffer; 5 μ l of α -[32 P]dCTP (3000 Ci/mmol); 0.002U DNase I; 10U DNA polymerase I; 100 ng probe.

The unincorporated nucleotides were removed on a G-50 column (NickTM column) equilibrated in TE (10 mM TrisHCl pH 8.0, 1 mM EDTA). Aliquots of the collected fractions were analyzed for the determination of the incorporated radioactivity (cpm).

A 162-bp promoter region from *S. solfataricus* genome was amplified with *Sso1078fw* and *Sso1078rv* primers, and used as a negative control for the EMSAs and as competitor in the competition assays. To determine if arsenite inhibited the binding of *TtSmtB* to the promoter DNA, various amounts of arsenic (0.5, 1, 2, and 4 mM) were added to the reaction mixtures. A typical reaction mixture (in a final volume of 15 μ l) contained 15,000 cpm (0.2 ng) of radiolabeled fragments, 1 μ g poly(dI-dC), and 3 μ g *TtSmtB* in binding buffer (25 mM Tris HCl, pH 8, 50 mM KCl, 10 mM MgCl₂, 1 mM dithiothreitol, 5% glycerol). The mixtures were incubated at 60°C for 20 min and run onto a nondenaturing 5% polyacrylamide gel (Bio-Rad) in 1X TBE at 80 V. The gel was dried and exposed to autoradiography. Alternatively, the gels were stained with SYBR Green.

4.3 Results

4.3.1 Transcriptional analysis of *TtsmtB* gene in *T. thermophilus* HB27 and its regulation by arsenic

To verify the transcription of the *TtsmtB* gene, a RT-PCR was carried out using as template DNaseI-treated total RNA, extracted from cells grown up to 0.5 OD_{600nm} and exposed to 8 mM NaAsO₂ and 12 mM KH₂AsO₄ for 0, 15, 30, 45, 60 min, and primers (*smtBfw* and *smtBrv*) annealing to the *TtsmtB* coding sequence in the +1 and +372 positions with respect to the putative translation initiation site. *TtsmtB* was found to be expressed under all the conditions tested (Fig. 1).



Fig. 1: Detection of *TtsmtB* transcript from untreated (lane 1) and treated- cells with arsenate for 15 (lane 2), 30 (lane 3), 45 (lane 4) and 60 (lane 5) minutes and arsenite for 15 (lane 6), 30 (lane 7), 45 (lane 8) and 60 (lane 9) minutes, by conventional RT-PCR. All RT-PCR products had the appropriate size (372 bp) on ethidium bromide stained agarose gel.

The genomic environment of *TtsmtB* is represented in Fig. 2, and includes a putative cluster from *TTC0351* to *TTC0355*, with the upstream and downstream genes (*TTC0350* and *TTC0356*) which have opposite orientation. *TTC0351*, encoding a cell wall endopeptidase of 388 amino acids, is separated from *TTC0352* by 134 bp. *TTC0352* and *TTC0355* encode two subunits of pyridoxal 5'-phosphate (PLP) synthase, an important enzyme in deoxyxylulose 5-phosphate (DXP)-independent pathway for *de novo* biosynthesis of PLP. *TTC0354*, separated by *TTC0353* by 33 bp, encodes a cation-transporting ATPase containing a HMA (heavy metal associated) motif.

In order to examine if a single transcript was formed with contiguous upstream and downstream genes, as suggested by bioinformatics, a RT-PCR analysis was performed; each primer set spanned the junctions between two adjacent genes. In particular, the reverse transcription reactions were obtained using alternatively: the primer *0352rv* (B in Fig. 3A), annealing to a sequence in the *TTC0352* gene at +138 from the putative start codon; the primer *0353rv* (D in Fig. 3A), annealing to a sequence in the *TTC0353* gene at +194; the primer *0354rv* (F in Fig. 3A), annealing

to a sequence in the *TTC0354* gene at +125; or the primer *0355rv* (H in Fig. 3A), annealing to a position +45 in the corresponding gene.

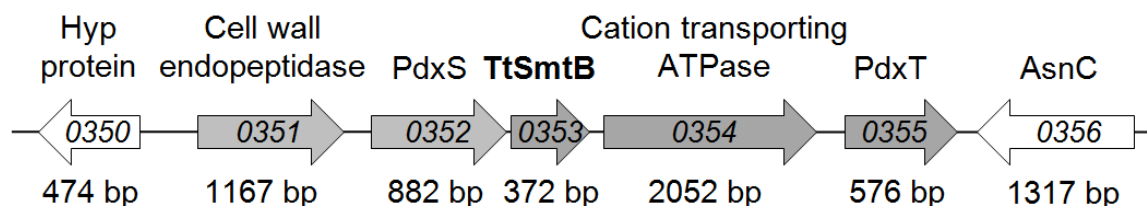


Fig. 2. Schematic representation of the genomic environment of *TtsmtB* gene. Arrowed boxes depict each ORF and the direction of transcription, with the locus tag given inside each box.

The four cDNAs obtained were then used in PCR reactions with primers *0351fw*, *0352fw*, *0353fw* and *0354fw* (A, C, E and G in Fig. 3A), respectively. From all primer pairs, fragments of the expected size were obtained (427 bp from AB; 221 bp from CD; 295 bp from EF; 635 bp from GH) (Fig. 3B). These results suggest that the four genes are co-transcribed as a polycistronic messenger. Transcription could initiate from a promoter putatively located in the 557 bp intergenic region between *TTC0350* and *TTC0351* genes. Investigation on the occurrence of similar gene associations in annotated bacterial genomes through Blast analysis showed an identical organization in diverse *T. thermophilus* strains (HB8, SG09) but also in *T. oshimai* and *scotoductus*, suggesting conservation in *Thermus* genome.

To verify if *TtsmtB* expression was arsenic dependent, its quantitative transcription was carried out in cells treated with arsenic compounds, in comparison with untreated cells, by preliminary qRT-PCR experiments. As shown in Fig. 4A, a two-fold increase of *TtsmtB* expression was observed in cell exposed to arsenate and arsenite. Arsenite induction was diminished after 60 minutes from the stress. To verify if also the expression of the downstream gene (*TTC0354*) was arsenic dependent, its quantitative transcription was carried out in cells treated with arsenic compounds in comparison with untreated cells. Preliminary qRT-PCR experiments showed an alternative regulation; no induction was observed in cells exposed to arsenate, whereas a strong induction was measured in arsenite-treated cells which reached a four-fold increase after 45 min of exposure to arsenite (Fig. 4B). Taken together these results highlight the presence of multiple arsenic responsive sequences, as the presence of arsenate responsive sequence internal to the operon, and strongly suggest an involvement of *TtsmtB* in arsenic detoxification. Further experiments are needed to evaluate the basis of the different regulation of the two clustered genes.

4.3.2 Cloning, expression and purification of TtSmtB

To demonstrate that TtSmtB was indeed a transcriptional regulator, the gene was cloned in pET28b(+) and pET30c(+) plasmids and expressed in *E. coli* BL21-CodonPlus(DE3)-RIL. All the attempts to obtain the protein in the soluble fraction or in pellet from pET30/*TtsmtB* transformed-cells were not successful. Whereas the recombinant His₆tag-fused protein was found in the soluble fraction.

After the determination of the optimal conditions (inducer concentration and exposure time) of TtSmtB expression, a standard expression protocol for the recombinant protein was developed. The pET28/*TtsmtB* transformed cells were grown up to 0.7 OD_{600nm} and then the expression was induced by addition of 0.5 mM IPTG, for additional 4 hours. The recombinant protein was purified to homogeneity through two

different steps: a heat treatment of the cell extract (70°C for 10 minutes), followed by affinity chromatography (Fig. 5A).

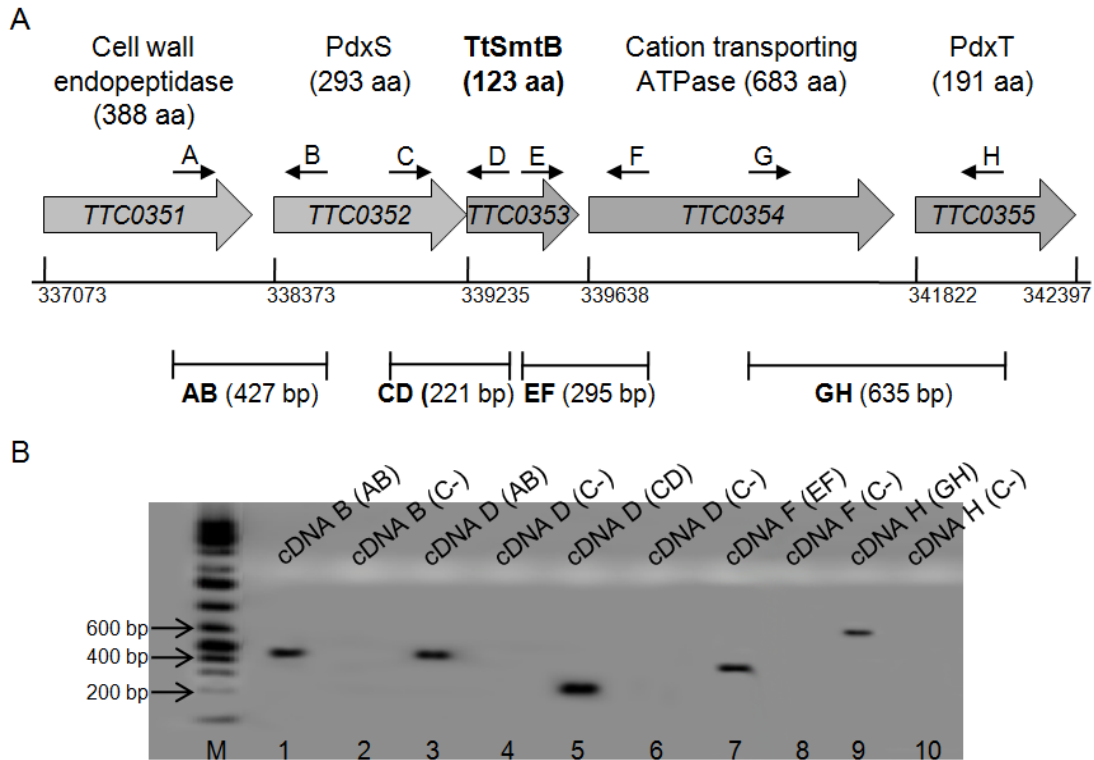


Fig. 3: (A) Schematic representation of the genomic environment of *TtsmtB*. Arrows above boxed ORFs depict annealing positions and orientation of the primers used. (B) Identification of the transcriptional unit. Agarose gel of RT-PCR products. All RT-PCR products had the expected size (AB: 427 bp, CD: 221 bp, EF: 295 bp, GH: 635 bp). M: ladder 1Kb plus; lanes 2, 4, 6, 8, 10: negative controls obtained using as template digested RNAs incubated without reverse transcriptase, lane 1: AB fragment from the 0352 cDNA; lane 3: AB fragment from the 0353 cDNA; lane 5: CD fragment from the 0353 cDNA; lane 7: EF fragment from the 0354 cDNA; lane 9: GH fragment from the 0355 cDNA.

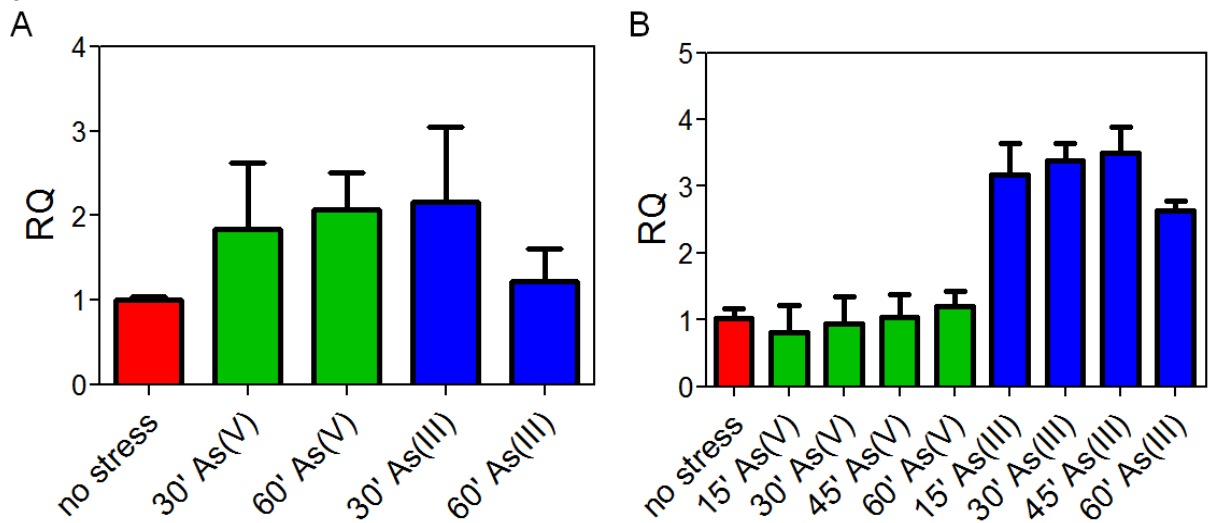


Fig. 4. Expression of *TtsmtB* and *TTC0354*, determined by qRT-PCR after stress by arsenate As(V) and arsenite As(III) at sub-inhibitory concentrations (12 mM and 8 mM) for 0, 15, 30, 45 and 60 minutes for *TTC0354*, and 30 and 60 minutes for *TtsmtB*. Error-bars represent SD of two independent experiments, each performed in triplicate.

From 100 ml culture about 3 mg of pure protein were obtained. The translated sequence encodes for a protein of 123 amino acids (13508.79 Da; pI 8.54). The expected molecular weight was confirmed by mass spectrometry. At all stages of purification the protein proved to be highly unstable and prone to degradation and aggregation. For this reason, 1 mM DTT, to maintain the cysteines in the reduced state, and 1X inhibitor cocktail, to prevent protease activity, were added to each passage.

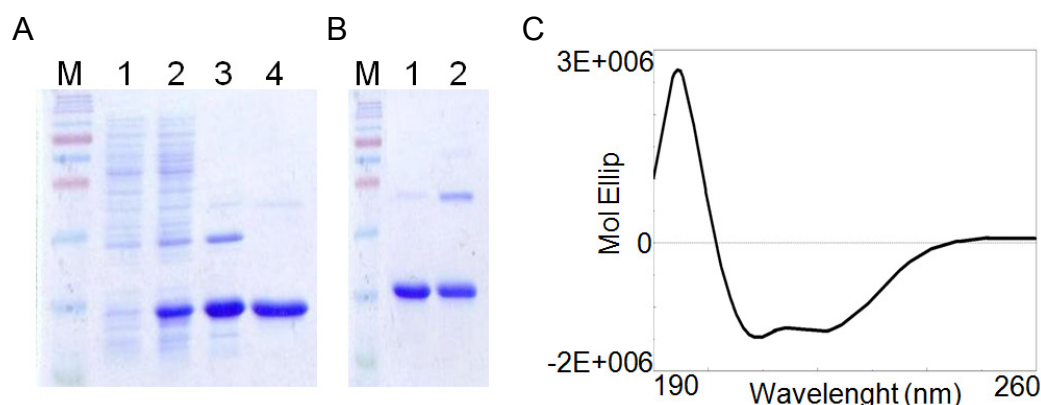


Fig. 5: (A) SDS-PAGE analysis of the purification of recombinant TtSmtB. Lane M, molecular mass marker; lane 1, cell extract from cells without pET28/*TtsmtB*; lane 2, crude extract; lane 3, heat-treated cell extract; lane 4, fraction from affinity chromatography. (B) Purified TtSmtB protein (8 µg) in the absence (lane 2) and presence (lane 1) of 2 mM 2-mercaptoethanol. (C) FAR-UV CD spectra of 15 µM TtSmtB.

4.3.3 Structural characterization of TtSmtB

Sequence analysis have shown that TtSmtb protein belongs to the ArsR/SmtB transcriptional regulator family. In fact, it has a conserved HTH DNA binding motif and a conserved ELCV(C/G)D metal binding box (Fig. 6). Alignments with homologues showed 50% identity with the structurally characterized SmtB transcriptional repressor from *Synechococcus* PCC7942. The high identity with the SmtB/ArsR members and the conservation of the cysteine residues involved in metal binding strongly suggest a key role for this protein in metal sensing. The secondary structure prediction showed a putative organization in 6 α -helices and 1 β -sheet, which is comparable to those found in ArsR/SmtB regulators (Fig. 6).

To assess the secondary structure of TtSmtB, CD spectra in the 190–260 nm region were recorded in 25 mM Na-P (pH 8.0) buffer at 25°C. As shown in Fig. 5C, the CD spectra of TtSmtB exhibited two negative peaks at 208 and 222 nm and one positive peak at 195 nm, indicative of a predominantly folded structure with a α - β content. The α -helical and beta content of the protein was estimated using Dichroweb 2.0 and corresponded to 18% and 51%, respectively, with the remaining percentage of random coil. These results are not in agreement with the structure predictions, but may indicate that the protein is unstable because it doesn't acquire its proper structure or loses it progressively. To determine the optimum, CD spectra were recorded at different pH, but no significant difference was observed.

Native gel filtration failed to reveal the quaternary structure of the recombinant protein, probably because of its aspecific interaction with the stationary phase. In fact, TtSmtB was eluted after 1.5 column volumes.

SDS-PAGE analysis showed that in the absence of reducing agents the recombinant TtSmtB reveals a band of size compatible with that of a dimeric protein (Fig. 5B).



Fig. 6: Multiple sequence alignment by Clustal W of TtSmtB (TTC0353) with SmtB/ArsR members. Proteins are: ArsR of *T. thermophilus* SG0, ArsR of *T. oshimai*, SmtB of *T. scotoductus*, SmtB of *D. radiodurans*, CadC of *C. perfringens*, SmtB of *Synechocystis*, ZiaR *Synechococcus* and ArsR of *E. coli*. The ELCVCD motif is highlighted by a box. Red filled arrows refer to the Cys62, and Cys64. Blue filled arrows refer to the putative residues included in the putative ligand binding site. The secondary structure elements of TtSmtB are depicted above the sequences.

4.3.4 Functional characterization of TtSmtB

To determine whether TtSmtB directly interacts with the putative promoters of *TTC0351/TTC0352/TTC0353/TTC0354/TTC0355* operon (*TtsmtB* operon) and *TtsmtB* gene, as well as the promoter region of the transcriptional unit containing the arsenate reductase gene (*TtarsC* operon) (described in the previous chapter), EMSA assays were performed. As shown in Fig. 7A, TtSmtb formed complexes with the fragments containing the putative promoters of the *TtsmtB* and *TtarsC* operons in a dose-dependent manner, and did not bind to the control DNA fragment amplified from the *Sso1078* promoter of *S. solfataricus*. As shown in Fig. 7B, TtSmtb also formed a complex with the fragment containing the putative regulatory sequences located upstream of the *TtsmtB* gene. To verify the activity of TtSmtB as a derepressor, as already shown for other family members, the effect of arsenite on the binding of TtSmtB to the promoter DNA were investigated by EMSA. The results

demonstrated that arsenite inhibited the formation of the *TtSmtB*-DNA complex in a dose-dependent manner and completely released *TtSmtB* from the DNA at a concentration of 4 mM (Fig. 7C).

4.4 Discussion

Resistance to both arsenite and arsenate is widely spread among both Gram⁻ and Gram⁺ bacteria, usually in the form of an *ars* operon.

The *ars* operons can be divided in two types with respect to gene organization. One type, found in the *S. aureus* pI258 plasmid and *E. coli* chromosomal *ars* operons, is composed of three arsenic resistance genes (*arsRBC*; encoding a regulatory repressor, a membrane transport pump and a cytosolic arsenate reductase, respectively); the second type, found in the *ars* operons of plasmids R773 and R46 of the Gram⁻ bacteria, consists of five genes (*arsRDABC*; where *ArsA* is an intracellular ATPase and *ArsD* is a metallochaperone) [11].

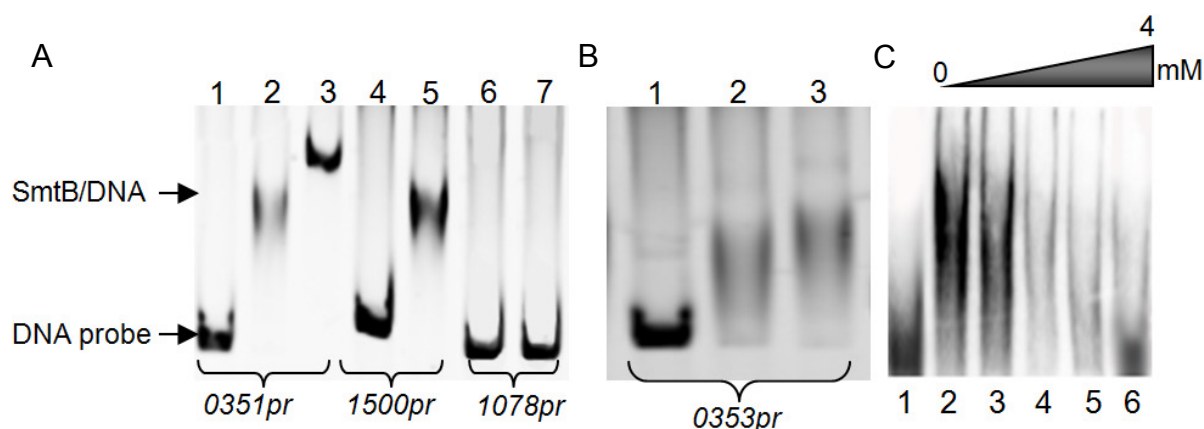


Fig. 7: (A) Binding of *TtSmtB* to the putative promoter regions of *TtsmtB* operon, *TtarsC* operon, and *Sso1078* fragment. EMSA results using 100 ng target DNA fragments: *TTC0351prom* (lane 1), *TTC1500prom* (lane 4), *Sso1078* (lane 6), incubated with 3 µg of purified *TtSmtB* (lanes 2, 5, 7, respectively) and 5 µg of purified *TtSmtB* (lanes 3). DNAs were stained by SYBR Green. (B) Binding of *TtSmtB* to the putative promoter region of *TtSmtB*. Lane 1: *TTC0353prom*, lane 2 and 3: *TTC0353prom* incubated with 3 µg and 5 µg of purified *TtSmtB*, respectively. (C) EMSA assay performed with 100 ng of *TtsmtB* fragment, 3 µg of *TtSmtB* and increasing concentrations of arsenite: 0.5 (lane 3), 1 (lane 4), 2 (lane 5), and 4 mM (lane 6).

T. thermophilus is a thermophilic microorganism which can be considered 'arsenic resistant' because it can grow at high arsenate and arsenite concentrations (20 and 12 mM of MIC, respectively). Compared with the *ars* operons in other bacteria, the *T. thermophilus* arsenic resistance genes are atypical in terms of genomic organization, because they are not organized in a single operon. In previous studies, we determined a high arsenic tolerance of *T. thermophilus* and in its genome we were able to find: a gene encoding a homologue of *ArsC* by *S. aureus*, four genes encoding putative transcriptional regulators, belonging to the *ArsR/SmtB* family, two genes encoding putative *ArsB* membrane transporters. No genes encoding putative *ArsA* or *ArsD* proteins were found.

The aim of this study was to characterize a putative *TtSmtB* transcriptional regulator and its role in the arsenic detoxification pathway. Among the putative transcriptional regulators found on the genome, *TTC0353* (*TtSmtB*) was chosen because it is the one that has the highest identity with the members of the *ArsR/SmtB* family

characterized so far. The genomic environment of *TtsmtB* includes a cluster from *TTC0351* to *TTC0355* genes, with opposing upstream and downstream genes (*TTC0350* and *TTC0356*). Results of end-point RT-PCR showed that the five genes are cotranscribed as an operon from a promoter upstream of the endopeptidase gene (*TTC0351*). This gene locus is highly conserved among other sequenced *Thermus* genomes (*T. thermophilus* HB8 and SG09, *T. oshimai* and *T. scotoductus*). Moreover, the predicted proteins encoded by the *TtsmtB*-flanking genes have little or no similarity to *ars* genes.

However, the possibility of another promoter located upstream of *TtsmtB* or *TTC0354* cannot be excluded; in fact, a putative bacterial promoter within these regions was identified by bioinformatic approaches. In particular, two putative promoters upstream of the operon, containing *TtsmtB*, and the *TtsmtB* gene, were identified; these promoters contain a 12-3-12 inverted repeat and two 6-2-6 inverted repeats, respectively. In addition, by qRT-PCR experiments, a two-fold increase of *TtsmtB* expression was observed in cells exposed to arsenate and arsenite. On the other hand, the same experiments performed on the immediately downstream gene (*TTC0354*), clustered with *TtsmtB* and encoding a putative cation transporting ATPase with 'heavy metal associated' motif, showed no induction in cells exposed to arsenate, whereas a four-fold increase was measured in arsenite-treated cells. The presence of internal promoters in an operon is not atypical, as it was also reported by Napolitano *et al*, who identified four internal promoters in a zinc-responsive operon from *Anabaena* PCC7120 cyanobacterium, tuned by several transcriptional regulators. Hence, a regulatory model compatible with these results has been proposed, where promoters in bacterial operons may be subject to a hierarchical regulation depending on their position in the operon.

The response to both arsenite and arsenate of *TtsmtB* is not atypical for ArsR/SmtB members; in fact, it has been reported that also ArsR2 of *Shewanella* increases its transcript levels in the presence of arsenite and in cells grown on arsenate [8]. Further testing is needed to determine whether *TtsmtB* transcription can be initiated from another arsenate and arsenite responsive promoter and whether the downstream gene of *TtsmtB* encodes a transporter which can extrude arsenite and to work as an ArsB pump. This hypothesis is also supported by qRT-PCR preliminary experiments carried out on the *TTC1447* transcript (annotated on the genome as putative arsenite efflux pump), which have shown that the gene may not be regulated by any of the two arsenic forms (arsenite and arsenate) (data not shown).

TtsmtB is a newly identified DNA-binding protein of 123 amino acids that shares significant sequence similarity with the Zn(II)/As(III)/Sb(III)-responsive repressors. *TtSmtB* was overexpressed in *E. coli* and purified to homogeneity. Gel filtration analysis could not confirm its putative dimeric structure, although suggested by structural data reported for previously characterized ArsR/SmtB members, because the protein is very unstable and prone to aggregation and precipitation. This is the main reason why its characterization proved so difficult. Indeed, the instability of the protein could be due to the presence of reactive cysteine residues.

The secondary structure predictions have shown a putative organization in 6 α -helices and 1 β -sheet, which is comparable to those found in ArsR/SmtB regulators. The members of this family are usually elongated dimers where each monomer is an α/β protein containing five α -helices and two β -sheets, as demonstrated for SmtB from *Synechococcus* [4]. Several common features of ArsR family proteins include their predicted size (12 to 16 KDa), helix-turn-helix domains, and several conserved cysteine residues [1]. *TtSmtB* has a conserved HTH DNA binding motif and the

majority of the conserved residues are located in the putative DNA and metal binding domains, but several residues in the α -6 helix (involved in dimerization) are also conserved.

A 12-3-12 inverted repeat in the putative *TTC0351* operator-promoter sequence was found, similar to 12-2-12 repeats found near or overlapping the transcriptional start site of the genes regulated by ArsR/SmtB proteins [1]. By EMSA assays it was possible to demonstrate that TtSmtB binds to this region. Two 6-2-6 inverted repeats in the putative promoter region of *TtsmtB* gene were found, suggesting the presence of a regulatory site in this region, as also demonstrated by the ability of TtSmtB to bind a fragment which includes these repeats. Since in both the analyzed sequences these inverted repeats overlap the putative -35 regions, the binding of TtSmtB to its operator may not inhibit the binding of RNA polymerase, but rather prevent the formation of basal transcriptional factors-promoter complex, ensuring repression of transcription.

In a previous study, we characterized an arsenate reductase from *T. thermophilus* with a key role in the arsenate reduction (Del Giudice *et al* 2013, submitted). In order to assess whether TtSmtB binds the putative promoter of the transcriptional unit containing *TtarsC*, an EMSA assay was performed. The results confirmed the ability of TtSmtB to bind this region.

ArsRs are known to bind and release DNA in the absence or presence of arsenite. A highly conserved ELCV(C/G)D motif was identified in members of this family and was proposed to contain residues directly involved in metal ion sensing [1]. TtSmtB shows a perfectly conserved ELCV(C/G)D motif. In TtSmtB, as also demonstrated for CadC and ArsR members, two of the three residues involved in the binding to the Zn(II) in SmtB of *Synechococcus*, Cys61 and Asp64, are conserved [4]. Cook *et al.* identified a metal binding site unique to SmtB (Asp104, His106, His117 and Glu120) [4], which is not conserved in all positions in the metalloregulated repressors. All four residues are conserved in TtSmtB.

Several cysteine residues in ArsR proteins are known to mediate the interactions with trivalent arsenic and antimony oxyanions. It has been shown that three cysteine residues (Cys32, Cys34 and Cys37) in *E. coli* ArsR bind to arsenic, but mutations in only two of them, Cys32 and Cys34, render the protein insensitive to inducers while preserving its ability to binding DNA [10]. In the alignment of TtSmtB with ArsRs, these cysteines correspond to conserved Cys62 and Cys64. Surprisingly, Cys61 of *Synechococcus* SmtB was found to be non-essential for Zn(II)-sensing *in vivo*, in contrast to Cys32 of ArsR [9,13]. This is the structural evidence that two distinct metal binding sites might exist within the family at least [1]. It seems likely that the TtSmtB protein contains both binding sites, suggesting its role in the sensing of arsenic/zinc or other metal ions, as reported for ZiaR of *Synechocystis* [12].

In conclusion, our study strongly suggests that TtSmtB is a specific regulator for arsenite, but that it might work in concert with other regulators to control the expression of *ars* genes. Further work, involving the demonstration of a direct interaction between TtSmtB and inducing compounds, mutational analysis of individual residues, and refinement of the purification process is required.

References

[1]Busenlehner LS, Pennella MA, Giedroc DP (2003) the SmtB/ArsR family of metalloregulatory transcriptional repressors: structural insights into prokaryotic metal resistance. FEMS Microbiol. Rev. 27:131-143.

- [2] Campbell DR, Chapman KE, Waldron KJ, Tottey S, Kendall S, Cavallaro G, Andreini C, Hinds J, Stoker NG, Robinson NJ, Cavet JS (2007). Mycobacterial cells have dual nickel-cobalt sensors: sequence relationships and metal sites of metal-responsive repressors are not congruent. *J Biol Chem*; 282:32298-32310.
- [3] Cavet JS, Graham AI, Meng W, Robinson NJ (2003). A cadmium-lead-sensing ArsR-SmtB repressor with novel sensory sites. Complementary metal discrimination by NmtR and CmtR in a common cytosol. *J Biol Chem*; 278:44560-44566.
- [4] Cook WJ, Kar SR, Taylor KB, Hall LM (1998). Crystal structure of the Cyanobacterial metallothionein repressor SmtB: a model for metalloregulatory proteins. *J Mol Biol*; 275:337-346.
- [5] Del Giudice I, Limauro D, Pedone E, Bartolucci S, Fiorentino G (2013). A novel arsenate reductase from the bacterium *Thermus thermophilus* HB27: its role in arsenic detoxification BBA Proteins and Proteomics. Submitted.
- [6] Liu T, Chen X, Ma Z, Shokes J, Hemmingsen L, Scott RA, Giedroc DP (2008) A Cu-sensing ArsR family metal sensor protein with a relaxed metal selectivity profile. *Biochemistry*; 47:10564-10575.
- [7] Morby AP, Turner JS, Huckle JW, Robinson NJ (1993). SmtB is a metal-dependent repressor of the cyanobacterial metallothionein gene *smtA*: identification of a Zn inhibited DNA-protein complex. *Nucleic Acids res*; 21:921-925.
- [8] Murphy JN, Saltikov CW (2009) The ArsR repressor mediates arsenite-dependent regulation of arsenate respiration and detoxification operons of *Shewanella* sp. Strain ANA-3. *J Bacteriol* 191:6722-6731.
- [9] Shi W, Dong J, Scott RA, Ksenzenko MY, Rosen BP (1996) The role of arsenic thiol interactions in metalloregulation of the *ars* operon. *J Biol Chem*; 271:9291-9297.
- [10] Shi W, Wu J, Rosen BP (1994) Identification of a putative metal binding site in a new family of metalloregulatory proteins *J. Biol. Chem.* 269:19826-19829.
- [11] Silver S, Phung LT (2005) A bacterial view of the periodic table: genes and proteins for toxic inorganic ions. *J Ind Microbiol Biotechnol*; 32: 587–605.
- [12] Thelwell C, Robinson NJ, Turner-Cavet JS (1998) An SmtB-like repressor for *Synechocystis* PCC6803 regulates a zinc exporter. *Proc Natl Acad Sci USA*; 95:10728-10733.
- [13] Turner JS, Glands PD, Samson ACR, Robinson NJ (1996) Zn²⁺ sensing by the cyanobacterial metallothionein repressor SmtB: different motifs mediate metal-induced protein-DNA dissociation. *Nucleic Acids Res*; 19:3714-3721.
- [14] VanZile ML, Chen X, Giedroc DP (2002). Structural characterization of distinct α 3N and α 5 metal sites in the cyanobacterial zinc sensor SmtB. *Biochemistry*; 41:9765-9775.
- [15] Wu J, Rosen BP (1991). The ArsR protein is a trans-acting regulatory protein. *Mol Microbiol*;5:1331-1336.

Chapter 5

GENERAL CONCLUSION

The environmental contamination with aromatic compounds and arsenic and their toxic effects on humans are a serious problem in many parts of the world. In Bangladesh, India, Vietnam and North America, for example, millions of people are still exposed to considerable levels of arsenic in drinking water [1,8]. Huge aromatic compound-contaminated sites exist worldwide; for example, in Germany more than 1400 sites contaminated by monoaromatic hydrocarbons, polycyclic aromatic hydrocarbons, phenolic and heterocyclic aromatic compounds, have been identified and only 20% of these areas has been considered for remediation [2].

Both arsenic and aromatic compounds are naturally present in the environment but human activities have contributed to their anomalous accumulation in the biosphere. From the above it is clear that the monitoring and remediation of waters and soils pollution, mainly near industrial zones or contaminated areas, is one of the major environmental objectives. Whole-cell biosensors have become a useful tool for monitoring environmental pollutants [9]. A microbial sensor system is a simple, specific, fast, cost-effective and sensitive method for measuring environmental pollutants. These bacterial sensors are usually genetically engineered to contain a reporter plasmid that carries a gene(s) and/or elements that respond to a target molecule; the response is indicated by expressing a reporter protein, such as the widely used green fluorescent protein, luciferase, β -galactosidase, or the recently reported carotenoid biosynthesis protein (CrtA) [10]. Very few attempts have been made to construct a whole-cell biosensor using hosts other than *E. coli* and, only mesophilic bacterial biosensors have been reported to date. The advantages of using extremophilic microorganisms are due to their ability to survive in harsh conditions, thus, they could be good candidates for the construction of more stable and stronger cellular or enzymatic biosensors. Extremophilic microorganisms represent a valuable resource in the development of novel biotechnological processes; in fact, most applications involving extremophiles are based on the use of their biomolecules, in particular their enzymes, which show important environmental benefits, such as the specific stability under extreme conditions. Among the extremophiles, thermophilic and hyperthermophilic microorganisms are probably the most studied organisms. The enzymes that have been isolated from these microorganisms are extremely thermostable and usually resistant to the action of chemical denaturants, detergents, chaotropic agents, organic solvents as well as to the exposure to extreme pH values. As a consequence, they can be used as biomolecular models for designing and constructing proteins with new properties of interest for industrial applications [4].

This work has been focused on the study of the strategies adopted by extremophilic microorganisms to colonize and survive in contaminated environments. In particular, the molecular mechanisms of response to environmental stresses have been thoroughly investigated. Furthermore, active and stable enzymes have been identified and characterized also in terms of their high thermal stability.

The first step of this research was the characterization of the detoxification pathway from aromatic compounds (principally benzaldehyde) and its regulation in the thermophilic *S. solfataricus* microorganism. In particular, the study has been focused on an archaeal MarR member, BldR2. Because of their function in multidrug resistance and tolerance to highly toxic compounds, MarR transcriptional regulators have been intensely characterized in bacteria and archaea with the aim of understanding the molecular mechanisms of regulated response to such a stress. BldR2 protein was overexpressed in *E. coli*, purified to homogeneity and characterized. The BldR2 dimer binds specifically to its own promoter region with micromolar affinity; the bound region contains an 8 bp pseudopalindromic sequence

separated by 3 bp from an 8 bp perfect palindrome. Since this site overlaps to the identified basal promoter elements, it can be proposed that the protein interferes with *bldR2* transcription by competing with basal factors. It is also demonstrated that BldR2 is able to bind benzaldehyde and salicylate, known to act as effectors of different MarR proteins, at micromolar concentration; salicylate and benzaldehyde were able to weaken the interaction of BldR2 with its own promoter. The *in vitro* binding results correlated with the *in vivo* induction of *bldR2* gene expression upon addition of aromatic drugs. Since it was also found that its gene expression was increased during the late-log growth phase, it was proposed that BldR2 could act *in vivo* to control regulatory mechanisms in detoxification from aromatic compounds complementary to that regulated by BldR, which mainly works in the exponential growth phase [5]. Furthermore, higher abundance of BldR2 in a later growth phase supports the hypothesis that *bldR2* expression levels could also be regulated by aromatic endogenous effectors deriving from aromatic catabolic pathways. Target genes for MarR regulators in archaea have only been identified in a few cases, but there is vast literature available from bacteria in which local transcriptional regulators are usually encoded upstream or downstream of the regulated genes. The analyses of the genomic environment of BldR2 revealed the adjacency of two divergent upstream *orfs*, *Sso1078* and *Sso1080*, encoding two components of an ABC-type antimicrobial peptide transport system. EMSA experiments demonstrated the ability of BldR2 to bind to sequences upstream of *Sso1078* and *Sso1080* with comparable affinities to those determined towards its own promoter. Since it was also found that BldR2 targets the *Sso2536* gene, encoding for an alcohol dehydrogenase involved in the detoxification of aromatic compounds, and the previously characterized MarR-like operon [6], a general role for BldR2 in the regulation of stress response genes in *S. solfataricus* has been proposed. The detoxification system could involve two MarR transcriptional regulators, BldR and BldR2, which bind to the promoter of MarR-like operon, encoding BldR and a drug export permease, and to the *adh* promoter, stimulating gene transcription, accumulation of the alcohol dehydrogenase and the enzyme-catalyzed conversion of the aldehydes to the less toxic alcohols. BldR2 binds also the promoter of an operon, encoding the components of antimicrobial peptide transport system, which could be involved in this pathway.

These findings could have interesting biotechnological applications, especially regarding the possibility of constructing whole-cell biosensors using the promoters responsive to aromatic compounds identified in this study; but further analyses are required to understand the overall mechanism of regulation by BldR2, which would include a genome-wide identification of target genes and natural ligands that the protein binds with a higher affinity.

The second step of this research was the characterization of the arsenic detoxification pathway and its regulation in another thermophilic microorganism: *T. thermophilus* HB27. This bacterium was shown to be able to survive in the presence of high concentrations of the most widespread arsenic forms in the environment (arsenite and arsenate). This ability has suggested the presence in the microorganism of genes encoding proteins involved in the mechanism of resistance to this metalloid. Arsenic resistance genes have been found on both the chromosomes and plasmids of various bacteria. These genes are usually arranged in an operon, the two most common forms of which contain either three (*arsRBC*) or five (*arsRDABC*) genes. In this PhD thesis, two arsenic resistance genes have been characterized in *T. thermophilus*, encoding an arsenate reductase (TtArsC) and a transcriptional regulator belonging to ArsR/SmtB family (TtSmtB). The genome of *T.*

thermophilus also contains two genes encoding putative arsenite efflux pumps, and other three genes encoding putative ArsR transcriptional regulators. Some microorganisms belonging to the *Thermus* genus are able to oxidize arsenite (by an arsenite oxidase) and to use the arsenate as the final electron acceptor in the anaerobic respiration (*arr* system) [6]. The genomic analysis of *T. thermophilus* HB27 revealed the absence of putative arsenite oxidase genes or *arr* genes involved in the arsenate respiration pathway. *T. thermophilus* HB27 only showed the genes of the *ars* system but with an unusual arrangement: in fact, they are not organized in a single operon. Transcriptional analysis of *TtarsC* showed that the gene is expressed and is the third of an operon also encoding proteins with no apparent metabolic relationship. Quantification of *TtarsC* transcription indicated low expression levels that increased up to four fold after cell exposure to arsenate. TtArsC was overexpressed in *E. coli*, purified to homogeneity and structurally-functionally characterized. It was proved that TtArsC is a monomeric enzyme able to convert arsenate to arsenite using electrons coming from the Tr-Trx system, with a catalytic mechanism involving the thiol group of the N-terminal Cys residue (Cys7) which performs a nucleophilic attack on the arsenate. The confirmation of the catalytic role of this cysteine was given by the inability of the C7S mutant to complete the Trx-coupled reaction cycle. Furthermore, TtArsC also retained a weak phosphohydrolase activity indicating its evolution from LMW PTPases and confirming the functional correlation with ArsC from *S. aureus* pl258. Intriguingly, TtArsC is the only characterized arsenate reductase from a Gram⁻ bacterium with such a dual catalytic activity (for example *E. coli* R773 ArsC has no measurable phosphatase activity) and it is one of the few described from Gram⁻ bacteria with a reduction mechanism characteristic of Gram⁺. Another unique feature of TtArsC is its thermostability. Its high T_m (91°C) and long activity half-life at high temperatures make it the most thermostable arsenate reductase described to date.

In order to determine how the arsenic resistance genes are regulated in *T. thermophilus*, TtSmtB protein was expressed, purified and preliminary characterized. Also in this case, the *TtsmtB* gene is transcribed in an operon of five genes which are not apparently metabolically correlated. A two-fold increase of *TtsmtB* expression was observed in cells exposed to arsenate and arsenite, but, the same experiments performed on the immediately downstream gene (*TTC0354*), encoding a putative cation transporting ATPase with 'heavy metal associated' motif, showed no induction in cells exposed to arsenate, whereas a four-fold increase was measured in arsenite-treated cells. The response to both arsenite and arsenate of *TtsmtB* is not atypical for ArsR/SmtB members [3]. Further testing is needed to determine whether *TtsmtB* and *TTC0354* transcription may be initiated from other arsenic responsive promoters.

The protein was able to bind its own putative promoter and the putative promoter of *TtarsC* operon. When added to the samples, the arsenite was able to alleviate the binding of TtSmtB to its own promoter, according to the presence in its primary sequence of a conserved arsenite binding site. Despite all attempts to increase the protein stability, it was not possible to perform additional characterizations of its DNA binding ability or to determine the quaternary structure. The fact that TtSmtB characterization is still preliminary has made it difficult to speculate on an arsenic detoxification strategy in *T. thermophilus*.

Taken together, our data reveal that the physiological function of TtArsC and TtSmtB could be a significant contribution to the arsenic resistance. Once entered into the cell, arsenate could increase *TtarsC* expression, probably through the coordinate action of TtSmtB and other unidentified transcriptional regulators. As the arsenate is

reduced by the enzyme, the even more toxic arsenite produced could be extruded by three putative arsenite permeases, two encoded by genes located on the chromosome and one encoded by a gene carried on the plasmid, which have not been functionally characterized yet; these transport genes could be regulated by TtSmtB.

Comprehensive knowledge on the molecular and genetic basis of detoxification, besides being stimulating from an evolutionary point of view, also represents an important starting point for developing efficient and selective arsenic bioremediation approaches. For example, a thermophilic arsenate reductase or a variant with improved characteristics obtained by directed evolution techniques, may be used in combination with materials capable of trapping the arsenite, as nanocrystalline magnetite [7].

References

- [1] Berg M, Tran HC, Nguyen TC, Pham HV, Schertenleib R, Giger W (2001) Arsenic contamination of groundwater and drinking water in Vietnam: a human health threat. *Environ Sci Technol*; 35:2621-2626.
- [2] Blum P, Sagner A, Tiehm A, Martus P, Wendel T, Grathwohl P (2011). Importance of heterocyclic aromatic compounds in monitored natural attenuation for coal tar contaminated aquifers: A review. *J Contam Hydrol*;126:181-194.
- [3] Butcher BG, Rawlings DE (2002) The divergent chromosomal *ars* operon of *Acidithiobacillus ferrooxidans* is regulated by an atypical ArsR protein. *Microbiology*; 148:3983-3992.
- [4] de Champdoré M, Staiano M, Rossi M, D' Auria S (2007). Proteins from extremophiles as stable tools for advanced biotechnological applications of high social interest. *J R Soc Interface*; 4:183–191.
- [5] Fiorentino G, Ronca R, Cannio R, Rossi M, Bartolucci S (2007). MarR-like transcriptional regulator involved in detoxification of aromatic compounds in *Sulfolobus solfataricus*. *J Bacteriol*; 189(20):7351-7360.
- [6] Gihring TM, Banfield JF, Arsenite oxidation and arsenate respiration by a new *Thermus* isolate, *FEMS Microbiol. Lett.* 204 (2001) 335-340.
- [7] Morin G, Wang Y, Ona-Nguema G, Juillot F, Calas G, Menguy N, Aubry E, Bargar JR, Brown GE Jr. (2009). EXAFS and HRTEM evidence for As(III)-containing surface precipitates on nanocrystalline magnetite: implications for As sequestration. *Langmuir*; 25 (16): 9119–9128.
- [8] Nordstrom DK (2002) Public health. Worldwide occurrences of arsenic in ground water. *Science*; 296:2143-2145.
- [9] Yagi K (2007) Applications of whole-cell bacterial sensors in biotechnology and environmental science. *Appl Microbiol Biotechnol*; 73:1251-1258.
- [10] Yoshida K, Inoue K, Takahashi Y, Ueda S, Isoda K, Yagi K, Maeda I (2008) Novel carotenoid-based biosensor for simple visual detection of arsenite: characterization and preliminary evaluation for environmental application. *Appl and Environ Microbiol*; 74:6730-6738.

List of talks and publications

Gabriella Fiorentino, Immacolata Del Giudice, Simonetta Bartolucci, Lorenzo Durante, Luigi Martino, Pompea Del Vecchio. (2011) 'Identification and physico-chemical characterization of BldR2 from *Sulfolobus solfataricus*, a novel archaeal member of the MarR transcription factor family'. *Biochemistry*, 50(31): 6607-6621.

Immacolata Del Giudice, Danila Limauro, Emilia Pedone, Simonetta Bartolucci, Gabriella Fiorentino. (2013). 'A novel arsenate reductase from the bacterium *Thermus thermophilus* HB27: its role in arsenic detoxification'. *BBA Proteins and Proteomics*, submitted.

Gabriella Fiorentino, Immacolata Del Giudice, Simonetta Bartolucci. 'Characterization of BldR2 from the archaeon *Sulfolobus solfataricus*: a new transcriptional regulator of the MarR family'. '**Proteins 2010**' Parma, Italy 8-10 April 2010.

Simonetta Bartolucci, Immacolata Del Giudice, Gabriella Fiorentino. 'BldR2, a new transcriptional regulator from the archaeon *Sulfolobus solfataricus* belonging to the MarR family'. **8th International Congress on Extremophiles**: 'Extremophiles 2010'. São Miguel, Azores, Portugal 12-16 September 2010.

Immacolata Del Giudice, Gabriella Fiorentino, Simonetta Bartolucci. Development of biological systems for the traceability of environmental pollutants. '**XI National Biotechnology Congress (CNBXI)**'. Varese, Italy 27-29 June 2012.

Identification and Physicochemical Characterization of BldR2 from *Sulfolobus solfataricus*, a Novel Archaeal Member of the MarR Transcription Factor Family

Gabriella Fiorentino,^{*,†} Immacolata Del Giudice,[†] Simonetta Bartolucci,[†] Lorenzo Durante,[‡] Luigi Martino,[§] and Pompea Del Vecchio^{*,‡}

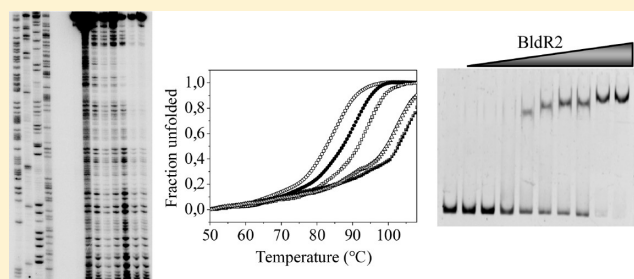
[†]Department of Structural and Functional Biology, University of Naples Federico II, Edificio 7, via Cinthia, 80126 Naples, Italy

[‡]Department of Chemistry "Paolo Corradini", University of Naples Federico II, via Cinthia, 80126 Naples, Italy

[§]Randall Division of Cell and Molecular Biophysics, King's College London, New Hunt's House, Guy's Campus, London SE1 1UL, U.K.

S Supporting Information

ABSTRACT: The multiple antibiotic resistance regulators (MarR) constitute a family of ligand-responsive transcriptional regulators abundantly distributed throughout the bacterial and archaeal domains. Here we describe the identification and characterization of BldR2, as a new member of this family, in the archaeon *Sulfolobus solfataricus* and report physiological, biochemical, and biophysical investigation of its stability and DNA binding ability. Transcriptional analysis revealed the upregulation of *BldR2* expression by aromatic compounds in the late-logarithmic growth phase and allowed the identification of *cis*-acting sequences. Our results suggest that BldR2 possesses in solution a dimeric structure and a high stability against both temperature and chemical denaturing agents; the protein binds site specifically to its own promoter, *Sso1082*, with a micromolar binding affinity at two palindromic sites overlapping TATA-BRE and the transcription start site. Benzaldehyde and salicylate, ligands of MarR members, are antagonists of binding of DNA by BldR2. Moreover, two single-point mutants of BldR2, R19A and A65S, properly designed for obtaining information about the dimerization and the DNA binding sites of the protein, have been produced and characterized. The results point out an involvement of BldR2 in the regulation of the stress response to aromatic compounds and point to arginine 19 as a key amino acid involved in protein dimerization, while the introduction of serine 65 increases the DNA affinity of the protein, making it comparable with those of other members of the MarR family.



Archaea, the third domain of life, are microorganisms adapted to grow in extreme environments with regard to temperature, pH, ionic strength, and high concentrations of detergents and organic solvents.¹ They are an evolutionary mosaic, being more similar to eukaryotes with respect to the macromolecular machinery and more similar to bacteria with respect to metabolic systems and genome organization.^{2–4} For example, with regard to the transcription machinery, in most cases, homologues of bacterial regulators function in the context of the archaeal basal transcriptional apparatus, which resembles that of Eukarya. Archaea possess a TATA box promoter sequence, a TATA box-binding protein (TBP), a homologue of the transcription factor TFIIB (TFB), and an RNA polymerase (RNAP) containing between 8 and 13 subunits.^{5,6}

Archaea have been of interest to many protein chemists over the years, as models for understanding the molecular basis of adaptation to extreme conditions. In the case of adaptations to extremes of pH, salinity, and pressure, it has been proven that membrane components and protective small molecules may play important roles.⁷ However, with regard to temperature

adaptations, the cellular components themselves, namely the proteins, have to achieve thermostability.^{8,9} Like all other living cells, archaea are also able to defend against subtle changes to environmental conditions; they own in their genomes finely regulated biochemical pathways for detoxification as well as different regulative sequences responsive to stress agents.^{10–13} The multiple antibiotic resistance regulators (MarR) constitute a family of ligand-responsive transcriptional regulators that are distributed throughout the bacterial and archaeal domains and include proteins critical for the control of virulence factor production, the response to oxidative stress, and the regulation of the catabolism of environmental aromatic compounds. They are also involved in the regulation of mechanisms of resistance to multiple antibiotics, organic solvents, household disinfectants, and pathogenic factors. MarR homologues are dimeric proteins that have a low level of sequence identity and a triangular shape; they

Received: February 4, 2011

Revised: June 28, 2011

Published: June 30, 2011

bind to their cognate palindromic or pseudopalindromic DNA as homodimers, resulting in either transcriptional repression or activation.¹⁴ The DNA binding domain of MarR proteins is a conserved winged helix–turn–helix motif¹⁵ with the two wings located at the corners of the triangle. Another common feature of MarR members is their ability to interact with specific ligands and, upon binding, to modulate DNA recognition.¹⁶ Crystal structures of several MarR regulators have been obtained, either as apoproteins, in complex with the cognate DNA, or with various effectors, greatly contributing to the elucidation of the mechanistic basis of DNA and/or ligand binding. However, the identification of key residues involved in binding as well as those contributing to protein stability and/or dimerization has been reported only in a few cases.^{17–19}

In the archaeal domain, the crystal structures of four transcription factors, ST1710 (or StEmrR) from *Sulfolobus tokodaii*,^{20,21} MTH313 from *Methanobacterium thermoautotrophicum*,²² PH1061 from *Pyrococcus horikoshii* OT3,²³ and BldR from *Sulfolobus solfataricus*,²⁴ have been determined. The overall structure of all of these proteins is typical of the MarR family, particularly for elements taking part in the DNA-binding domain, while the most important differences are found in the dimerization domain. The structures of MTH313 and ST1710 complexed with salicylate as the ligand revealed conservation of the ligand binding pocket.^{17,22} Furthermore, it has been proposed that the ability to act as activators or repressors could not be related to a particular DNA binding mechanism, but rather to the position of the binding site on the target DNA.²⁴ Comparison of the thermophilic StEmrR with mesophilic MarR showed an increase in the number of salt bridges on the protein surface predicted to be important in increasing the thermostability.²¹

Some of us identified BldR as part of an operon-like structure conserved in most archaea. Functional characterization demonstrated that BldR is a transcriptional activator involved in the detoxification of aromatic compounds.^{25,26} The *S. solfataricus* genome contains an additional ORF, *Sso1082*, that encodes another putative MarR protein of 154 amino acids that is 35% identical in sequence to BldR of the same organism.

In this work, we have performed a physiological and biochemical analysis of BldR2 with the aim of enhancing our knowledge of the role of MarR proteins in the archaeal domain. Our results showed that the gene is transcriptionally regulated. Furthermore, BldR2 is a homodimer that binds specifically to its own promoter in a region that overlaps with the sequences recognized by the basal transcription machinery. In the presence of benzaldehyde and salicylate, ligands of MarR members, BldR2 dissociates from its promoter. In this study, we also report interesting results in terms of conformational stability and DNA binding properties of BldR2 through biophysical and biochemical measurements. Moreover, to provide further insights into thermal stability and DNA binding molecular mechanisms of BldR2, we used guided mutagenesis based on the structure of the close relatives BldR²⁴ and ST1710²¹ and identified two residues, Arg19, possibly involved in dimer stabilization, and Ala65, located in the DNA binding domain. BldR has in the same position a serine residue directly involved in DNA binding.²⁴ These two residues, Arg19 and Ala65, were substituted in two different mutants with Ala and Ser, respectively, and a complete characterization was conducted in parallel with wild-type BldR2.

MATERIALS AND METHODS

***S. solfataricus* Cultivation and Preparation of Genomic DNA and Total RNA.** *S. solfataricus* P2 was grown aerobically at 82 °C in 100 mL of medium described by Brock supplemented with 0.1% (w/v) yeast extract and 0.1% (w/v) casamino acids²⁷ and buffered at pH 3.5. In some cases, benzaldehyde, sodium salicylate, and benzyl alcohol were added to final concentrations of 1, 0.35, and 4 mM, respectively, after dilution of an exponentially growing culture up to an A_{600} of 0.08 optical density (OD) unit. Cells were grown to ~0.3 and ~0.7 OD₆₀₀ unit, corresponding to midlogarithmic and late-logarithmic phases, respectively, and harvested by centrifugation at 4000g for 10 min. Genomic DNA and total RNAs were prepared following reported procedures.²⁸

In Vivo Response to Aromatic Compounds. RNAs (10 µg) extracted from cells grown under different conditions were electrophoretically separated in a 1.5% agarose gel containing 10% formaldehyde and transferred to nylon filters (GE Healthcare). Hybridization was conducted as described by Cannio et al.,²⁸ using the *Sso1082* and *Sso1352* genes²⁵ and rRNA 16S as probes. The experiments were performed in duplicate. Signals were visualized by autoradiography and quantified with a densitometric analysis using a Personal Fx phosphorimager and Quantity One (Bio-Rad).

Primer Extension Analysis of the Transcription Start Site. To determine the first transcribed nucleotide, total RNA extracted from cells grown in the presence or absence of benzaldehyde and harvested at 0.3 OD₆₀₀ unit was subjected to primer extension analysis as described by Limauro et al.,²⁹ using primer *Sso1082+100Rv* (5'-GGC CTA TTT GCT CAA GAG CC-3'), annealing from position 100 bp in the *Sso1082* gene. The same primer was used to produce a sequence ladder by using the f-Mol DNA cycle sequencing system (Promega), according to the manufacturer's instructions, to locate the products on a 6% urea–polyacrylamide gel.

Heterologous Expression of *Sso1082* and Purification of the Recombinant Protein. The gene encoding *Sso1082* from *S. solfataricus* P2 was amplified by polymerase chain reaction (PCR) from genomic DNA, using *Pfx* DNA polymerase. Two different upstream primers both containing the *NdeI* restriction site (underlined) (*Sso1082up*, 5'-GGA TTT TGT GAG TTCATATGA TG-3'; *Sso1082Fw*, 5'-GTT AGA TAT CTA CAT ATG ATA TTA GC-3') were designed on the basis of two different putative translation start sites. In particular, *Sso1082Ssoup* anneals to the ATG annotated on the genome³⁰ while *Sso1082Fw* anneals to another putative start codon located 45 nucleotides downstream and deduced from both a transcriptional analysis³¹ and our primer extension. The common downstream primer *Sso1082 Rev* (5'-GCT TTA AGA CTC GAG TAG TTA GG-3') introduces a stop codon and the *XhoI* restriction site (underlined in the sequence). Amplified fragments were purified, digested with appropriate restriction enzymes, and cloned in the pET30a *NdeI*- and *XhoI*-modified vector, generating pET30*Sso1082long* and pET30*Sso1082*, respectively. The sequences of the two cloned fragments were shown to be identical to those available on the *S. solfataricus* P2 genome (<http://www-archbac.u-psud.fr/projects/sulfolobus/>).

Escherichia coli BL21-CodonPlus (DE3) cells (Stratagene) transformed with pET30*Sso1082long* and pET30*Sso1082* were used for recombinant protein expression.

Cells transformed with pET30Sso1082 were grown in 1 L of Luria-Bertani medium containing kanamycin (100 $\mu\text{g}/\text{mL}$) and chloramphenicol (33 $\mu\text{g}/\text{mL}$). When the culture reached an A_{600} of 0.5 OD unit, protein expression was induced by the addition of 0.5 mM IPTG and the bacterial culture grown for an additional 6 h. Cells were harvested by centrifugation, and pellets were lysed by sonication in 30 mL of lysis buffer [50 mM Tris-HCl (pH 8.0) and 1 mM phenylmethanesulfonyl fluoride] in an ultrasonic liquid processor (Heat system Ultrasonic Inc.). The lysate was centrifuged at 30000g for 60 min (SW41 rotor, Beckman). The supernatant was heated to 80 °C for 10 min, and denatured proteins were precipitated by centrifugation at 20000g for 20 min at 4 °C.

The supernatant was loaded onto a heparin column (5 mL, HiTrap heparin, GE Healthcare) equilibrated in 50 mM Tris-HCl (pH 8.0) (buffer A) connected to an AKTA Explorer system (GE Healthcare). After a washing step with buffer A, elution was conducted with 40 mL of a KCl gradient (0 to 0.8 M). Fractions containing the BldR2 protein were pooled, concentrated, dialyzed, and loaded onto a Superdex 75 column (26 cm \times 60 cm, GE Healthcare) equilibrated in 50 mM Tris-HCl (pH 8.0) and 0.2 M KCl (buffer B) at a flow rate of 2 mL/min. Fractions were collected and analyzed by SDS-PAGE to detect the BldR2 protein. These fractions were pooled, concentrated by ultrafiltration using a YM10 membrane (Millipore), dialyzed against buffer A, and stored at 4 °C.

Construction, Expression, and Purification of R19A and A65S Mutants. Single-point mutations in the *BldR2* gene were produced with the Quick Change Site-Directed Mutagenesis Kit (Stratagene) that utilizes *PfuUltra* high-fidelity DNA polymerase and primers complementary to the coding and noncoding template sequences.

To generate the R19A and A65S mutants, the forward primers (*FwSso1082R19A*, 5'-TTGGGATCTAATAACTGGACTGACAGCAAAGATAAATAAGGATACAGATAAG-3'; *FwSso1082A65S*, 5'-GCTGAAAAATATATGCTGACAAAGTCGGGATTAAGTAGCATCATT-3') with their complementary reverse primers were used (underlined letters indicate the base pair mismatch) and reactions performed according to the manufacturer's instructions. The mutagenesis products were transformed into XL-1 Blue Cells. Single colonies were selected on LB plates containing kanamycin, and isolated plasmidic DNAs were sequenced at Eurofins MWG Operon.

Plasmids pET30R19A and pET30A65S containing the desired mutations were used to transform BL21 Codon plus (DE3) competent cells. The best growth conditions for gene expression were determined both for cells transformed with pET30R19A and for cells transformed with pET30A65S: growth to an OD_{600} of ~ 0.5 and ~ 0.8 in LB medium supplemented with kanamycin (50 $\mu\text{g}/\text{mL}$) and chloramphenicol (33 $\mu\text{g}/\text{mL}$) at 37 °C, respectively, followed by induction for 6 h with 0.5 mM IPTG.

The purification of the mutant proteins was conducted in a manner similar to that already described for the wild-type BldR2 protein.

Computational Methods. To establish similarities among the sequences of proteins of interest and the sequences of the SwissProt Data Bank, computational analysis was performed at <http://www-arch-bac.u-psud.fr/projects/sulfolobus/> or <http://cmr.jcvi.org/cgi-bin/CMR/GenomePage.cgi?org=ntss02>, providing access to the genome of *S. solfataricus* P2.

The multiple-sequence alignment was obtained using T-COFFEE.³² The three-dimensional (3D) model of BldR2 has

been built with EsyPred3D using the BldR 3D structure as a template.³³

Analytical Methods for Protein Characterization and Sample Preparation. Protein concentrations were determined by the method of Bradford, using BSA as the standard.³⁴ Protein homogeneity was estimated by SDS-PAGE [12.5% (w/v) gels]. To determine the native molecular mass of the proteins, the purified proteins at different concentrations (0.25, 0.5, 1.5, and 3.0 mg/mL) were applied in a volume of 100 μL to an analytical Superdex PC75 column (3.2 cm \times 30 cm) connected to an AKTA Explorer system (GE Healthcare) alternatively equilibrated with buffer B, with 20 mM sodium phosphate (pH 7.5) and 0.2 M KCl, or with the same buffer used for EMSA (see below), at a flow rate of 0.04 mL/min. The column was calibrated in the different buffers using a set of gel filtration markers (low range, GE Healthcare), including bovine serum albumin (67.0 kDa), ovalbumin (43.0 kDa), chymotrypsinogen A (25.0 kDa), and RNase A (13.7 kDa). The molecular mass of the protein and mutants was also determined using electrospray mass spectra recorded on a Bio-Q triple quadrupole instrument (Micromass, Thermo Finnigan, San Jose, CA).

Protein solutions for spectroscopic analyses were prepared in a 20 mM sodium phosphate buffer (pH 7.5), and the concentration was determined by UV spectra using a theoretical, sequence-based extinction coefficient of 22920 $\text{M}^{-1} \text{cm}^{-1}$ calculated at 280 nm for the dimeric protein.³⁵ A commercial 8 M GuHCl solution from Sigma was used as a denaturant solvent. Protein solutions for CD and fluorescence measurements were exhaustively dialyzed by using Spectra Por membranes (molecular weights of 15000–17000) against buffer solution at 4 °C. The water used for buffer and sample solutions was doubly distilled. The pH was measured at 25 °C with a Radiometer model PHM93 pH meter.

Samples for GuHCl-induced denaturation experiments were prepared with increasing amounts of denaturing agent. Each sample was mixed by vortexing and incubated at 4 °C for 1 day. Longer incubation times produced identical spectroscopic signals.

Cloning of *Sso1082* Promoter Regions. Two different regions upstream of the ORF *Sso1082* were amplified by PCR amplification on *S. solfataricus* P2 genomic DNA: the first of 230 bp was obtained using the primer pair *Sso1082fw-130* (5'-ATT AGG ATA TAG ATC TCG TTT ACG A-3') and *Sso1082+100Rv*. The *Sso1082prFw* primer anneals starting at position -130 with respect to the transcription initiation site as determined by primer extension analysis.

A second smaller region of 164 bp was amplified with the pair *Sso1082fw-130* (see above) and *Sso1082+34Rv* (5'-CCCAAAC-TTCTGAGTACTTTGTAG-3') annealing from position +34 in the *Sso1082* gene.

The fragments were cloned in the pGEM T-easy vector (Promega) and TopoTA (Invitrogen) to give *Sso1082L* (large) and *Sso1082S* (small), respectively. The insertion and correct sequence of the PCR products were verified by DNA sequencing.

DNA Binding. Binding of BldR2, R19A, and A65S to the putative regulatory region *Sso1082L* was measured by an electrophoretic mobility shift assay (EMSA) using a biotin-labeled PCR fragment amplified from the pGEM T-easy vector using the *Sso1082Fw-130* and the *Sso1082+100Rv* primers (see above). The amplified DNA fragment was labeled at the 3'-OH end with the Biotin 3' End DNA labeling kit (Thermo Scientific), according to the manufacturer's protocol.

EMSA reaction mixtures were set up in a final volume of 15 μ L and contained 5 nM DNA, 1 μ g of poly(dI-dC), and varying amounts of proteins in binding buffer [25 mM Tris-HCl (pH 8), 50 mM KCl, 10 mM MgCl₂, 1 mM dithiothreitol, and 5% glycerol]. The mixtures were incubated at 60 °C for 20 min and run onto a nondenaturing 5% polyacrylamide gel (Bio-Rad) in 1 \times TBE at 80 V. The probes were transferred onto a positively charged nylon membrane (Hybond-XL, GE Healthcare, Uppsala, Sweden) with a blotting apparatus (Bio-Rad, Hercules, CA) and then detected with the Chemiluminescent Nucleic Acid Detection Module Kit (Thermo Scientific) according to the kit protocol.

To determine the dissociation constants of BldR2, R19A, and A65S with respect to the *Sso1082* promoter, the *Sso1082S* regulatory region (150 nM) was incubated with increasing amounts of the purified proteins and the complexes were separated as described above; after electrophoresis, gels were directly stained with SYBR green (nucleic acid gel stain, Invitrogen). At least two independent experiments were performed in duplicate. In particular, protein concentrations (in dimer units) ranged from 1.0 to 90 μ M for BldR2 and R19A and from 0.05 to 1.5 μ M for A65S.

Densitometric data were obtained with Quantity One (Bio-Rad) and manipulated to calculate the fractional complex formation (that is the ratio between the density of the retarded band and the total density, reported in percent). These values were analyzed by fitting the binding isotherm to the Hill equation in GraphPad Prism 5.0.

To analyze the effect of benzaldehyde and salicylate on the properties of BldR2, R19A, and A65S binding to the *Sso1082S* promoter, EMSAs were performed via preincubation of the proteins, at concentrations similar to their apparent K_d values, with 3, 5, 10, 20, 30, 50, and 100 mM sodium salicylate or benzaldehyde. Gels were processed and visualized as described above.

DNase I Footprinting. A probe containing the promoter region of the *Sso1082* gene was produced by PCR using a combination of the *Sso1082* *ftR* primer (5'-CGAATTCGCCCTT-GGGTTTGAAG-3') designed on the basis of the *Sso1082* promoter (bold) and pTopo sequences, respectively, starting at +34 bp from the transcription initiation site, and *Sso1082*-184 (5'-CCATATTTATAATCTCTACA-3') as a second primer; the latter was labeled at the 5' end with T4 polynucleotide kinase and [γ -³²P]ATP. The labeled PCR product of 231 bp (~40 nM) was incubated with 5–10 μ g of pure BldR2, R19A, and A65S at 60 °C in binding buffer (see above) and digested with 1 unit of DNase I (Ambion) for 1 min at 37 °C. Subsequent steps were performed as described by Fiorentino et al.²⁵ Labeled primer was as also used to generate a dideoxynucleotide sequence ladder with the Promega f-Mol DNA sequencing system using *Sso1082S*-Topo (see above) as the template and following the manufacturer's instructions.

Circular Dichroism and Fluorescence Measurements. CD spectra were recorded with a Jasco J-715 spectropolarimeter equipped with a Peltier-type temperature control system (model PTC-348WI). The molar ellipticity per mean residue, $[\theta]$ in degrees square centimeters per decimole, was calculated from the equation $[\theta] = ([\theta]_{\text{obs}} \text{mrw}) / (10lC)$, where $[\theta]_{\text{obs}}$ is the ellipticity measured in degrees, mrw is the mean residue molecular weight (117.6) for protein BldR2, C is the protein concentration in grams per milliliter, and l is the optical path length of the cell in centimeters. Cells with path lengths of 0.1 and 1 cm were used in the far-UV and near-UV regions, respectively. CD spectra were recorded with a time constant of 4 s, a 2 nm bandwidth, and a

scan rate of 20 nm/min; the signal was averaged over at least three scans and baseline corrected by subtraction of a buffer spectrum. Spectra were analyzed for secondary structure amount according to the CDSSTR method³⁶ using Dichroweb.³⁷ The GuHCl-induced denaturation curves, at a fixed constant temperature of 25 °C, were obtained by recording the CD signal at 230 nm for the samples containing increasing amounts of GuHCl. Finally, the thermal unfolding curves were recorded in the temperature mode, by following the change in the CD signal at 222 nm with a scan rate of 1.0 °C/min. Fluorescence measurements were performed with a JASCO FP-750 apparatus equipped with a circulating water bath to keep the cell holders at a constant temperature of 25 °C. The denaturant-induced unfolding curves were obtained by recording changes in both fluorescence intensity and fluorescence maximal wavelength as a function of GuHCl concentration. The excitation wavelength was set to 290 nm, and the experiments were performed by using a 1 cm sealed cell and a 5 nm emission slit width and corrected for the background signal. The change in fluorescence intensity at 336 nm was recorded to monitor the unfolding transition. The protein concentration was kept constant at 2.4 μ M.

Analysis of the Denaturant-Induced Unfolding Transitions. For comparison of the denaturant-induced unfolding curves obtained in both the CD and fluorescence experiments, the curves were normalized reporting the fraction of unfolded protein (f_U) as a function of the concentration of the denaturing agent.

Thermodynamic parameters for the denaturant-induced unfolding were determined by analyzing the transition curves on the basis of a simple two-state model for dimeric proteins. In the equilibrium, only folded dimeric protein N_2 and unfolded monomers U exist. At any point in the denaturation reaction, the equilibrium constant K_U was calculated according to the model

$$K_U = [U]^2/[N_2] = 2P_t(f_U^2)/(1 - f_U) \quad (1)$$

in which P_t is the total molar concentration of protein monomers. The midpoint of the unfolding transition, C_m , was calculated using the equation

$$C_m = [RT \ln(P_t) + \Delta G_U(H_2O)]/m$$

The unfolding Gibbs energies were calculated using the relation

$$\Delta G_U = -RT \ln K_U \quad (2)$$

where R is the gas constant and T the absolute temperature. The linear dependence of the Gibbs energy of unfolding on the denaturant concentration is given by

$$\Delta G_U = \Delta G_U(H_2O) + m[D] \quad (3)$$

where $\Delta G_U(H_2O)$ is the extrapolated Gibbs energy of unfolding in the absence of denaturant³⁸ and m is the cooperativity parameter.³⁹ Values of $\Delta G_U(H_2O)$ were obtained directly by fitting the unfolding curves to eqs 1–3.

RESULTS

Transcriptional Analysis of *Sso1082*. Northern blot experiments were performed to verify the transcription of the *Sso1082* gene in two growth phases and to analyze transcription in cells grown in the presence of aromatic compounds, benzaldehyde, benzyl alcohol, and salicylate; these compounds were already demonstrated to inhibit cell growth at 1.5, 4, and 0.5 mM,

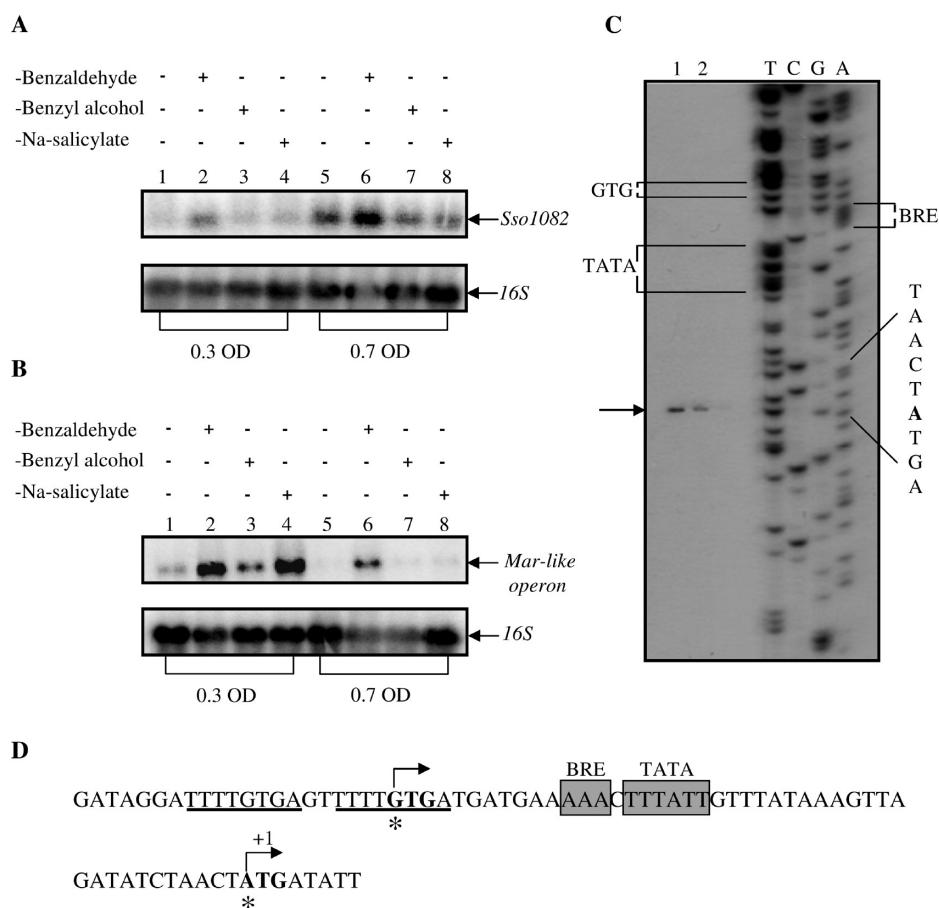


Figure 1. Transcriptional analysis of the *Sso1082* gene. Northern blot analysis of *Sso1082* (A) and *Mar-like operon* (B) mRNAs. Total RNA was prepared from *S. solfataricus* cells grown in the presence of different aromatic compounds and harvested in exponential (lanes 1–4) and stationary (lanes 5–8) growth phases: lanes 1 and 5, untreated control cells; lanes 2 and 6, cells grown in the presence of 1 mM benzaldehyde; lanes 3 and 7, cells grown in the presence of 4 mM benzyl alcohol; lanes 4 and 8, cells grown in the presence of 0.35 mM salicylate. The filters were probed with the *Sso1082* (A) and *Sso1352* (B) genes. Amounts of the mRNAs were normalized to 16S rRNA. The experiments were performed in duplicate. (C) Primer extension analysis of the *Sso1082* promoter region. Total RNA was prepared from cells grown in the presence (lane 1) or absence (lane 2) of benzaldehyde and harvested in the exponential growth phase. Primer-extended products were separated by electrophoresis under denaturing conditions alongside sequencing reactions with the same primer. (D) Promoter sequence analysis. The mapped transcription/translation start site (+1) is highlighted in bold; TBP and TFB binding sites are boxed. The initiation codon as annotated on the *S. solfataricus* genome is in bold and TSS as determined by Wurtzel are indicated by an asterisk.

respectively, and to affect *MarR-like operon* transcription.²⁵ The *BldR2* expression pattern was compared to that of the *Mar-like operon* and the 16S rRNA. *BldR2* mRNA revealed a single hybridization band under all the conditions tested with a molecular transcript of ~400 bp, which is slightly lower than that deduced from the gene sequence (462 bp) but is in accordance with a monocistronic transcript (Figure 1A). A densitometric analysis (Figure S1 of the Supporting Information) revealed that the level of *Sso1082* mRNA was ~3-fold higher in cells grown in the late exponential growth phase in comparison with that of cells grown in the early exponential growth phase (Figure 1A and Figure S1 of the Supporting Information, lanes 1 and 5).

Furthermore, the level of the transcript increased ~2-fold in cells grown in the presence of 1 mM benzaldehyde in comparison with nontreated cells in both growth phases (Figure 1A and Figure S1 of the Supporting Information, lanes 1, 2, 5, and 6). A weak induction could be observed when challenging cells with 4 mM benzyl alcohol (Figure 1A and Figure S1 of the Supporting Information, lanes 1–3 and 5–7) or 0.35 mM salicylate (Figure 1A and Figure S1 of the Supporting Information, lanes 1–4 and 5–8). The amounts of total cellular RNA were

comparable in all the experiments, which could be judged by hybridization of the same filters with the 16S rRNA gene. Taken together, these results are evidence that the level of *Sso1082* expression increases in a later growth stage with respect to that of the *Mar-like operon* and responds to stress by aromatic drugs (Figure 1B).

To determine the transcription start site of *Sso1082*, a primer extension analysis was undertaken on RNAs prepared from cells grown in the presence or absence of benzaldehyde. Unexpectedly, the results presented in Figure 1C reveal that the transcription start site corresponds to an adenosine located 45 bp downstream of the GTG start codon annotated on the *S. solfataricus* P2 genome.³⁰ Interestingly, a recent report on the *S. solfataricus* transcriptome has revealed that *Sso1082* transcription starts at two major positions: one is located in position –1 relative to the beginning of the ORF annotated on the genome and a second that coincides with the one we mapped and overlaps a putative ATG start codon; the two transcripts would produce proteins in the same frame differing for 15 amino acids at the N-terminus.³¹ It was also found that the extent of transcription from the downstream start site was >10-fold higher.³¹

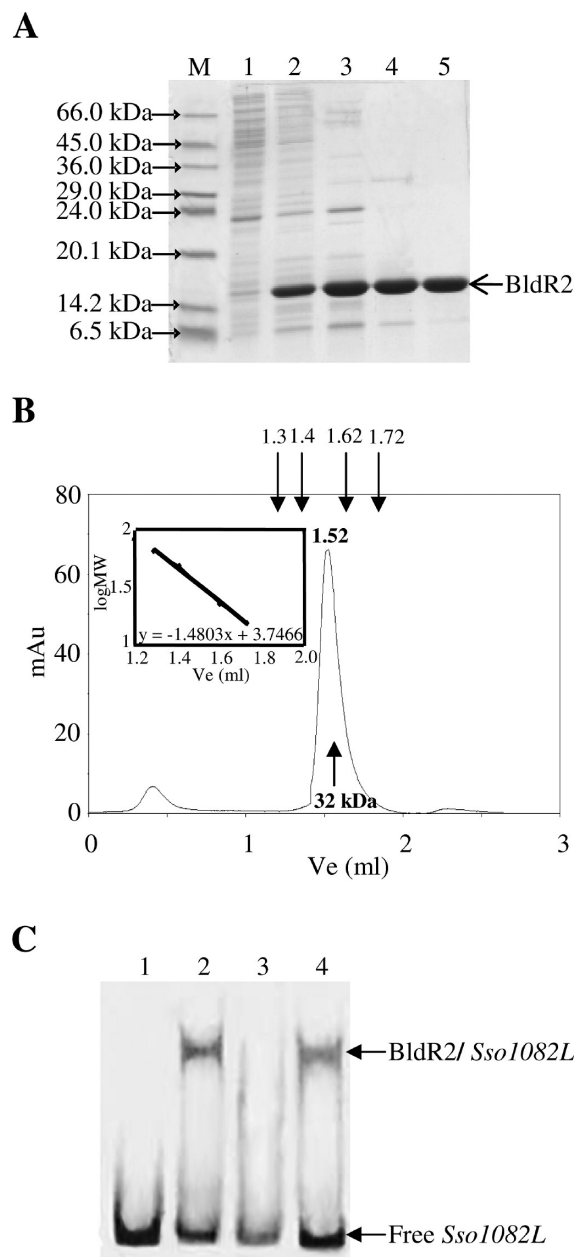


Figure 2. Analysis of recombinant BldR2. (A) SDS–PAGE of the purification steps of recombinant BldR2: lane M, molecular mass markers; lane 1, crude extract; lane 2, heat-treated cell extract; lane 3, fraction from heparin chromatography; lane 4, fraction from exclusion molecular chromatography; lane 5, purified BldR2. (B) Elution profile of purified BldR2 from gel filtration on a Superdex PC75 column. Recombinant BldR2 is eluted at 1.52 mL corresponding to a molecular mass of 32 kDa. Arrows indicate the elution volumes of the protein standards in the relative calibration of the column. (C) EMSA. *Sso1082L* (5 nM) and BldR2 incubated in the presence of specific and nonspecific competitors: lane 1, labeled DNA fragment; lane 2, DNA probe and BldR2 (15 μ M); lane 3, *Sso1082L*, BldR2, and 1250 nM nonlabeled *Sso1082L* fragment; lane 4, *Sso1082L*, BldR2, and 1250 nM nonlabeled *gfp* fragment.

However, under all of our growth conditions, we detected only the shorter transcript. It is tempting to speculate that under our experimental conditions transcription and translation start sites could coincide, giving rise to a single transcript, and in vivo, a

mature protein could be also translated starting from a Met residue located downstream from that found in the genome annotation.

The hypothesis is strengthened by the fact that the transcriptome study proved that leaderless translation is the preferred strategy in *S. solfataricus*.³¹ On the basis of the position of the downstream transcription start site, it was possible to identify TATA box and BRE sequences perfectly matching with the consensus, located downstream of the predicted GTG initiation codon (Figure 1D). The results depicted in Figure 1C confirm the relative increase in the level of mRNA upon benzaldehyde induction.

Heterologous Expression and Characterization of BldR2.

Both putative *Sso1082* genes predicted from genome annotation²⁶ and from the transcriptional analysis³¹ were cloned in pET30 and expressed in *E. coli*; interestingly, BldR2 was expressed in *E. coli* as a soluble protein only when the corresponding gene was cloned from the downstream start codon. All of our attempts to produce a soluble protein from the gene sequence as annotated in the *S. solfataricus* genome failed;³⁰ the protein purified from inclusion bodies was also found to be very sensitive to protease degradation at its N-terminus. Edman sequencing revealed, in fact, the absence of the first 13 amino acids (data not shown).

Hence, in the following description, we refer to BldR2 as the smaller protein translated from the downstream transcription and translation start site, and the reported characterization has been performed on this protein. BldR2 was purified to homogeneity (Figure 2A), taking advantage of its intrinsic properties, including its thermostability, its putative DNA binding capability, and its small size. Active fractions were selected on the basis of their ability to shift their own promoter region, *Sso1082L*, spanning positions –130 to 100 relative to its transcription and translation initiation codon in an EMSA (see also below). From 1 L of *E. coli* culture, it was possible to obtain up to 20 mg of pure BldR2. The molecular mass of the recombinant BldR2 as determined by MS analysis was 16348 Da, in agreement with the corresponding theoretical value, and the same applies to the mutants. The quaternary structure was assessed via analytical gel filtration and revealed a homodimeric structure both at different protein concentrations and in different buffers (Figure 2B and Figure S2 of the Supporting Information), a result consistent with other MarR homologues.¹⁵

Specific Binding of BldR2 to Its Promoter Region. Because the majority of MarR transcription factors are autoregulators that bind site-specifically to their promoters, we assessed whether BldR2 had such a capacity. An EMSA confirmed that BldR2 was able to bind to this region (Figure 2C, lane 2) and revealed the specificity of the interaction; in fact, a *gfp* gene fragment²⁶ at a 250-fold molar excess could not compete for BldR2 binding (Figure 2C, lane 4), whereas an unlabeled specific competitor abolished gel retardation when added at the same ratio (Figure 2C, lane 3). Hence, this result shows that BldR2 specifically recognizes its own promoter.

Structural and Functional Characterization of BldR2 and Its Mutants. A multiple-sequence alignment of BldR2 with archaeal MarR members whose structure is known and comparison with bacterial representatives identified four identical residues, namely, Leu29, Leu74, Thr97, and Gly100 (Figure 3A). Leu29, Thr97, and Gly100 are located in helices α 1 and α 5 in the dimerization domain, while Leu74 is in helix α 4 in the DNA binding domain.^{13,24} Residues Glu75, Arg89, and Glu95,

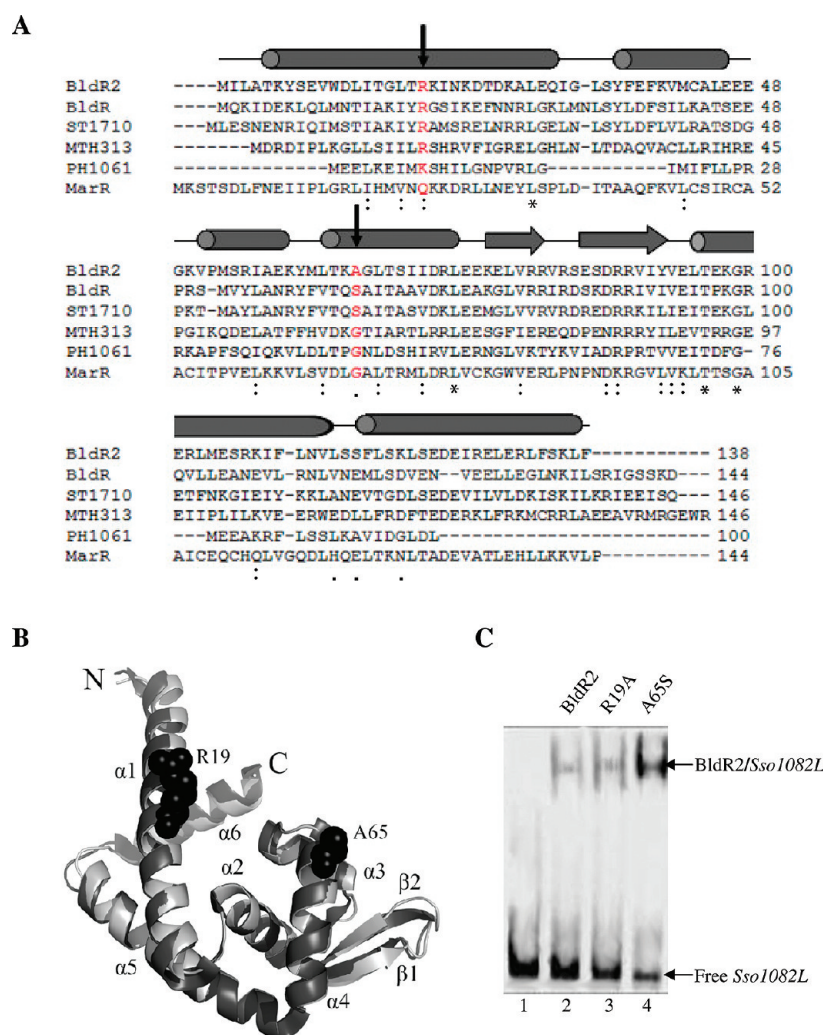


Figure 3. (A) Multiple-sequence alignment of BldR2 (*Sso1082*) with characterized MarR members. Proteins are BldR (*Sso1352*) from *S. solfataricus*, ST1710 from *S. tokodaii*, MTH313 from *M. thermoautotrophicum*, PH1061 from *P. horikoshii*, and MarR from *E. coli*. The secondary structure elements of BldR2 are depicted above the sequences. The mutations introduced by site-directed mutagenesis are highlighted by arrows. (B) Superposition of the 3D structure of the monomer of BldR (dark gray, Protein Data Bank entry 3F3X) and the 3D model of BldR2 (light gray). The mutated residues are shown as gray spheres. (C) Binding of purified BldR2, R19A, and A65S to the *Sso1082L* promoter region: lane 1, *Sso1082L*; lane 2, *Sso1082L* and 15 μM BldR2; lane 3, *Sso1082L* and 15 μM R19A; lane 4, *Sso1082L* and 15 μM A65S.

occurring in strand β3 (the wing motif), are identical only in the archaeal domain. Hence, on the basis of sequence analysis, BldR2 is expected to share a function similar to that of the other family members.

Sequence analysis of BldR2 from *S. solfataricus*, in comparison with its closest homologue, BldR, from the same organism, provided further information for the design of mutants that could help shed light on the structural properties and functions of BldR2. On the basis of the homology sequence shared between BldR and BldR2, we used the known 3D structure of BldR as a reference model for BldR2 (Figure 3B).²⁴ By also looking at the sequence alignment of Figure 3A, we found that some of the key residues important for dimer stabilization and DNA interaction in BldR are present in BldR2. Among these, the arginine in position 19 has been found to be conserved in known archaeal representatives. In most members with known structure, the Arg residue is in helix α1 of the dimerization domain and is located at the dimer interface where it contributes to dimer stability.^{15,24} Furthermore, BldR2 has an alanine in position 65 instead of a

serine that is conserved in BldR and ST1710 and is involved in DNA binding.^{17,24} To investigate the contribution of these two amino acids to the DNA binding properties as well as the stability of BldR2, two mutants were generated. In the first, the arginine at position 19 was substituted with an alanine; in the second, the alanine at position 65 was substituted with a serine. For each of the single mutants, we performed a characterization in parallel with the wild-type protein. The mutants were overexpressed in *E. coli* and purified using the same procedure described for the wild-type enzyme. Native gel filtration revealed their ability to form dimers, suggesting that the mutations introduced did not alter the monomer–monomer interactions (Figure S2 of the Supporting Information). To verify whether the mutations affected the DNA binding capability of BldR2, we employed EMSAs using the *Sso1082L* promoter as the target DNA. Figure 3C shows that both mutants retained their activity, again indicating that the mutations were not disruptive. Interestingly, an increased intensity of the shifted band could be observed when the promoter region was incubated with identical amounts

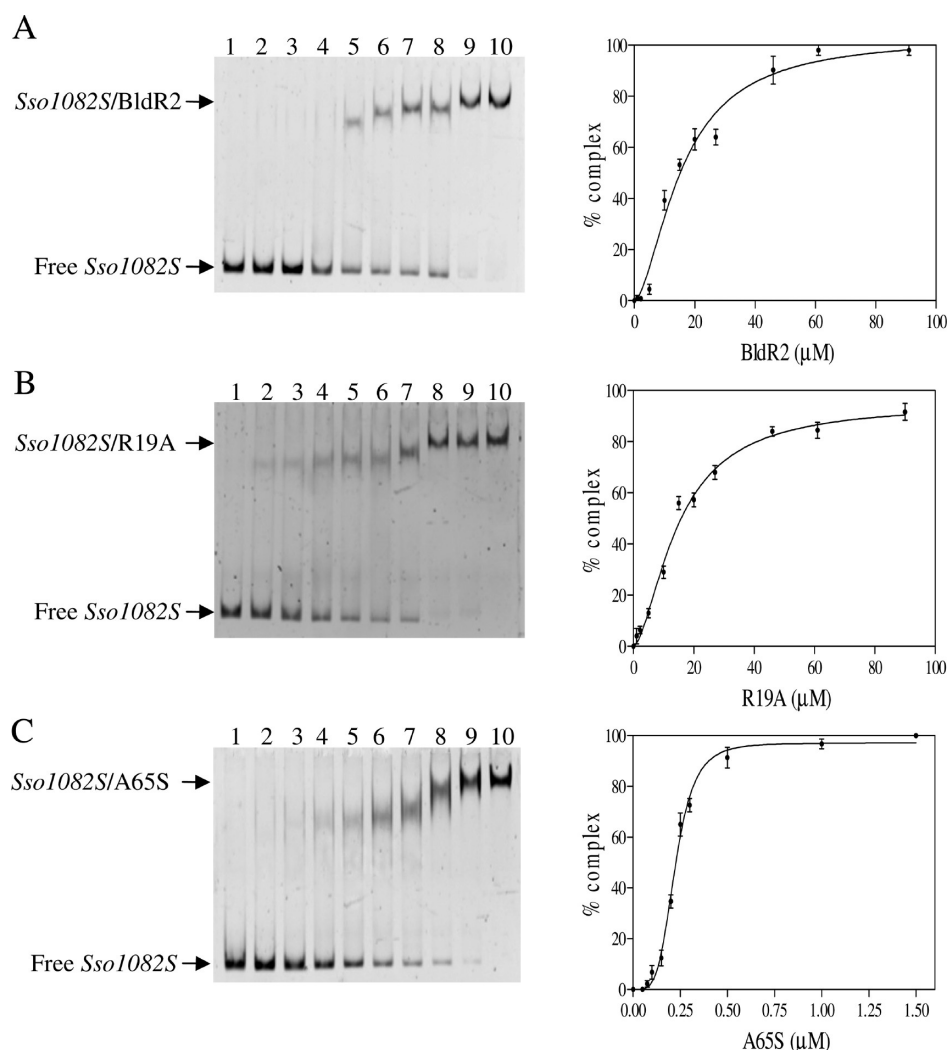


Figure 4. Binding of BldR2, A65S, and R19A to the *Sso1082* promoter assessed by an EMSA. (A) *Sso1082S* (150 nM) titrated with increasing concentrations of BldR2: lane 1, labeled DNA fragment; lanes 2–10, DNA probe incubated with BldR2 at concentrations ranging from 1.5 to 91 μM . (B) *Sso1082S* (150 nM) titrated with increasing concentrations of R19A: lane 1, labeled DNA fragment; lanes 2–9, *Sso1082S* incubated with R19A at concentrations ranging from 1.5 to 91 μM . (C) *Sso1082S* (150 nM) titrated with increasing concentrations of A65S: lane 1, labeled DNA fragment; lanes 2–9, *Sso1082pr* incubated with A65S at concentrations ranging from 0.05 to 1.5 μM . Densitometric data from EMSA obtained as described in Materials and Methods are plotted vs the concentration of each protein (right of each panel). Error bars represent the standard deviation for each point derived from four experiments.

of the A65S mutant (Figure 3C, lane 4) with respect to the wild-type protein (lane 2) and R19A (lane 3).

DNA Interaction Studies. To gain insight into the biological role of BldR2, we analyzed in more detail its DNA binding compared to that of its mutants, testing by an EMSA their binding affinity for an *Sso1082* promoter region, spanning positions –184 to +34 relative to the transcription and translation initiation codon (*Sso1082S*).

Titration with increasing amounts of the BldR2 dimer indicated that the protein binds its own control region in a concentration-dependent manner; furthermore, at saturating concentrations, the protein determined a shift with decreased mobility, suggesting either that more binding sites with different affinities could exist in the DNA sequence analyzed or that multiple dimers could associate to the cognate DNA. The profile obtained by fitting densitometric data to a binding curve with a Hill slope gave an overall apparent equilibrium dissociation

Table 1. Dissociation Constants Calculated by EMSA for BldR2, R19A, and A65S with the *Sso1082* Promoter

	BldR2	R19A	A65S
<i>Sso1082S</i>	$15.8 \pm 5.2 \mu\text{M}$	$16.3 \pm 4.5 \mu\text{M}$	$0.22 \pm 0.02 \mu\text{M}$

constant (K_d) of 15.8 μM (Figure 4A and Table 1) and provided a Hill coefficient of 1.7. A comparable binding pattern (Figure 4B) with a similar global affinity ($K_d = 16.3 \mu\text{M}$) and slope (1.6) was observed for the R19A mutant, while mutation of A65 to serine significantly changed the DNA binding properties; in fact, formation of a complex with A65S was seen using much lower protein concentrations (Figure 4C), and fitted binding data indicated an ~ 70 -fold increased affinity for the target promoter ($K_d = 0.22 \mu\text{M}$) and a slope of 4.3. These results clearly suggested that A65S binds with high affinity and via a different mechanism with respect to those of BldR2 and R19A.

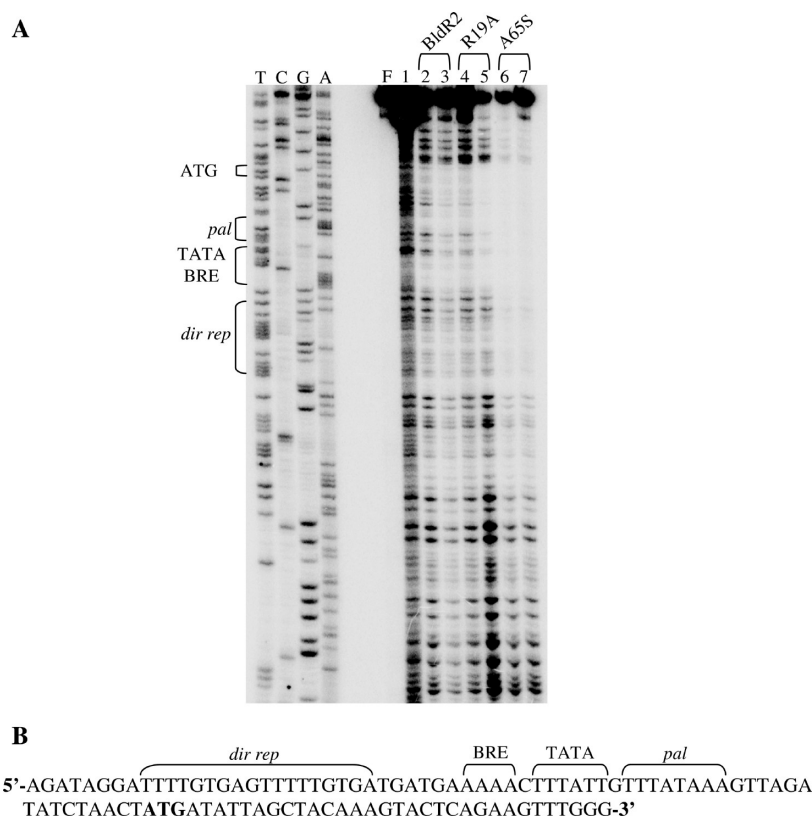


Figure 5. (A) Wild-type and mutant binding sites of the *Sso1082* promoter. DNase I footprinting analysis was performed at the nontemplate strand using 0.0 μ g (lane 1), 5 μ g ($\sim 15 \mu$ M, lanes 2, 4, and 6), and 10 μ g (lanes 3, 5, and 7) of purified BldR2, R19A, and A65S. DNA fragments were analyzed in parallel with a sequencing reaction (relative lanes are indicated by the corresponding nucleotide positions on the top) by denaturing gel electrophoresis. (B) Positions of the footprints indicated on the nucleotide sequences relative to the transcription and translation start site.

With the aim of precisely positioning the binding sites of BldR2 at its own promoter, a DNase I footprinting analysis was undertaken. As shown in Figure 5A, BldR2, R19A, and A65S protect a region of ~ 40 bp extending from the pseudopalindromic TATA-BRE sequences to the ATG start codon and containing a 8 bp TTTATAAA palindromic site (Figure 5B). The extent of protection suggests that more than one BldR2 dimer could associate to target DNA. Furthermore, these results prove the presence of regulatory sites in the BldR2 promoter that could function for its *in vivo* autoregulation. Interestingly, the A65S mutant gave an even more extended footprint; in fact, it protected a further region of 17 bp located more upstream of the TATA box, containing the direct repeat TTTTGTGAgTTTGTGTA (Figure 5B). This evidence could imply that the introduction of a serine enhances the DNA binding affinity by modifying the recognition sequence.

To analyze the DNA binding behavior of BldR2 upon addition of putative phenolic ligands, we performed EMSAs in which formation of the complex of BldR2 and the mutants with the *Sso1082S* promoter was tested after preincubation of the proteins with increasing concentrations of salicylate and benzaldehyde. The first ligand was chosen because it was demonstrated to be the negative effector of several MarR homologues,^{17,22,41} while the second was known to interact positively with the BldR factor in *S. solfataricus*.²⁵

The results are shown in Figure 6A: an inhibition of complex formation starts at ~ 10 mM salicylate (lane 5), while the benzaldehyde was slightly less effective (Figure 6B). In fact, release of

BldR2 from its own promoter could be observed from ~ 30 mM benzaldehyde (lane 6).

DNA binding by the mutant A65S was also affected in the presence of the tested ligands (Figure 6C,D) to a similar extent; in addition, we observed a further band above that of the free DNA, which was interpreted as a partial dissociation of the DNA–protein complex upon ligand interaction.

Taken together, these data indicate that salicylate and benzaldehyde are low-affinity ligands of BldR2 that associate with the protein to attenuate DNA binding.

Conformational Stability of Dimeric BldR2. According to its hyperthermophilic origin, BldR2 has proved to be highly thermostable and resistant to the GuHCl denaturing action as supported by detailed investigation of its conformational stability by means of CD and fluorescence measurements. Figure 7 shows the CD spectra of BldR2 in the far-UV region at 25 $^{\circ}$ C (spectrum a) and 105 $^{\circ}$ C (spectrum b). From the latter spectrum, it appears that BldR2 is still folded at that high temperature; in fact, it retains the characteristic minima at 222 and 208 nm with a small decrease in the CD band intensities. The complete disappearance of the canonical CD bands occurs when the protein is incubated for 24 h with 7 M GuHCl (spectrum c). The analysis of the CD spectrum at 25 $^{\circ}$ C, performed using Dichroweb,³⁷ suggested that BldR2 contains 66% α -helix and 9% β -sheet; these values closely resemble those obtained from the X-ray structure of the homologous protein, BldR.^{24,25} The near-UV CD spectrum of BldR2 is reported in the inset of Figure 7. It also suggested that the protein possesses a well-defined conformation at 25 $^{\circ}$ C in buffer

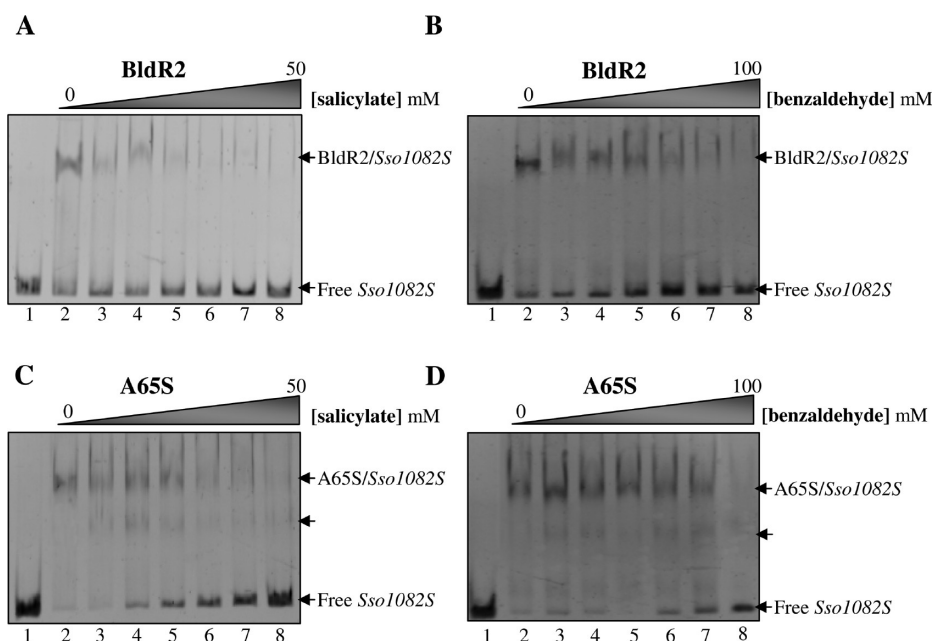


Figure 6. Formation of BldR2–*Sso1082S* and A65S–*Sso1082S* complexes in the presence of salicylate and benzaldehyde. (A and B) *Sso1082S* (150 nM) and BldR2 (16 μ M) titrated with increasing concentrations of salicylate (A) and benzaldehyde (B): lane 1, DNA fragment; lane 2, *Sso1082S* incubated with BldR2; lanes 3–8, *Sso1082S* and BldR2 incubated with 3, 5, 10, 20, 30, and 50 mM salicylate (A) and 3, 10, 20, 30, 50, and 100 mM benzaldehyde (B), respectively. (C and D) *Sso1082S* (150 nM) and A65S (0.2 μ M) titrated with increasing concentrations of salicylate (C) and benzaldehyde (D): lane 1, DNA fragment; lane 2, *Sso1082S* incubated with A65S; lanes 3–8, *Sso1082S* and A65S incubated with 3, 5, 10, 20, 30, and 50 mM salicylate (C) and 3, 10, 20, 30, 50, and 100 mM benzaldehyde (D), respectively.

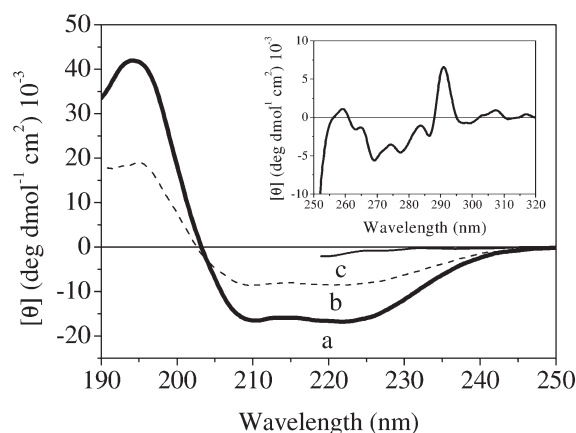


Figure 7. (A) Far-UV CD spectra of BldR2. Spectra were recorded at 25 °C (a, —), 105 °C (b, ---), and 25 °C after incubation for 24 h with 7 M GuHCl (c, ···). The inset shows the near-UV CD spectrum of BldR2 (—). The protein concentrations used to acquire the spectra are 2.4 and 20 μ M in the far- and near-UV regions, respectively.

solution. The CD spectra of both mutants A65S and R19A were found to be very similar to those of BldR2, suggesting that the secondary and tertiary structure are not affected by the mutations (Figure S3 of the Supporting Information).

Thermal Unfolding. Changes in the far-UV CD signal at 222 nm have been used to follow the thermal unfolding of BldR2 in 20 mM sodium phosphate buffer (pH 7.5). Because of the very high thermal stability, it was not possible to obtain a complete thermal unfolding curve as the transition is not yet finished at 105 °C [see the filled squares (■) in Figure 8], and the instrumental setup cannot work at

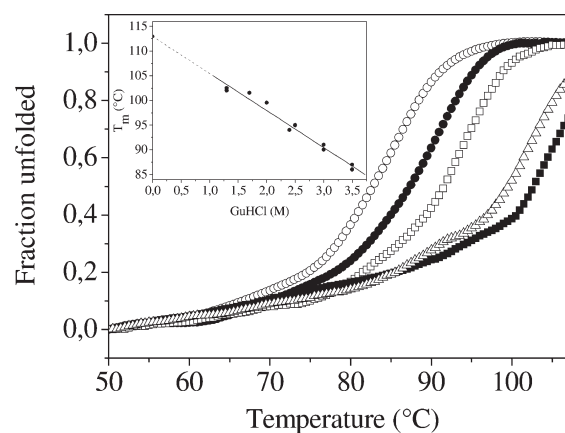


Figure 8. Thermal denaturation curves of BldR2 at a fixed concentration of 2.4 μ M, in 20 mM sodium phosphate (pH 7.5) in the presence of 0 (■), 1.3 (△), 2.4 (□), 3 (●) and 3.5 M GuHCl (○). The curves were obtained by recording the changes in the molar ellipticity at 222 nm as a function of temperature. The inset shows the dependence of the melting temperature T_m on GuHCl concentration; via linear extrapolation, a value for T_m in the absence of the denaturant agent was estimated.

temperatures higher than 110 °C. To obtain complete thermal denaturation curves, solutions of BldR2 at a fixed final concentration of 2.4 μ M were incubated with increasing GuHCl concentrations to progressively destabilize the native state. The corresponding thermal denaturation curves are collected in Figure 8, and the denaturation temperatures obtained were as follows: 102 °C at 1.3 M GuHCl, 96 °C at 2.4 M GuHCl, 91 °C at 3 M GuHCl, and 85 °C at 3.5 M GuHCl. Via linear extrapolation of these numbers up to 0 M GuHCl, the estimated denaturation temperature of BldR2

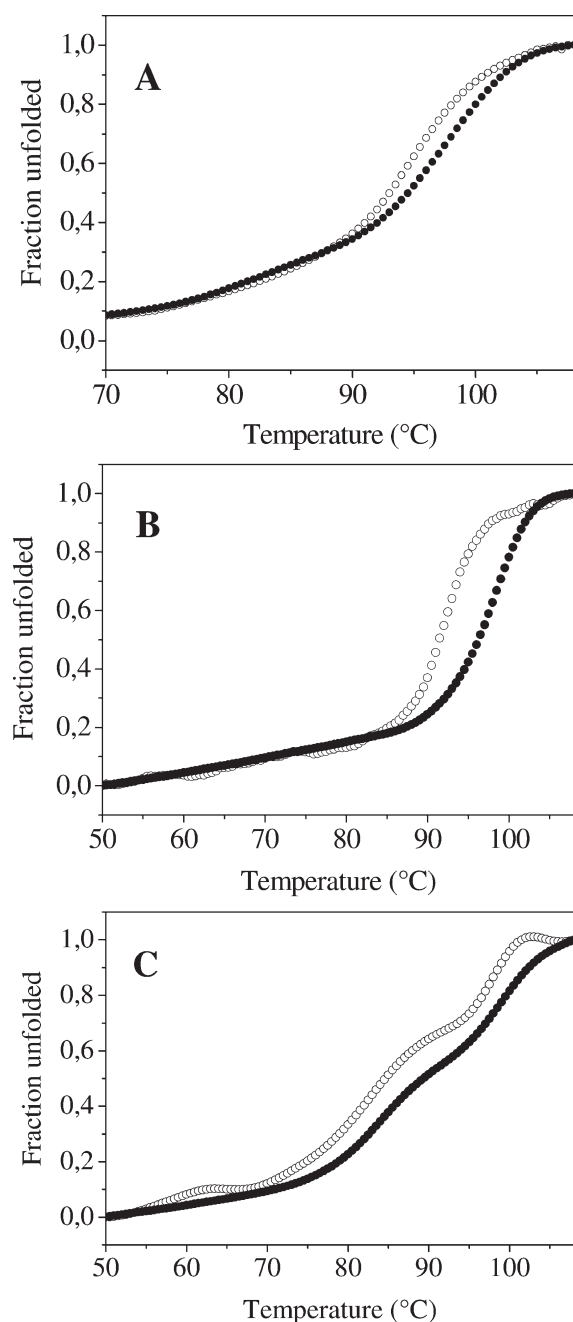


Figure 9. (A) Concentration dependence of BldR2 thermal denaturation, in the presence of 2.4 M GuHCl. Protein concentrations were 1.5 (○) and 9 μM (●), and the T_m shifted from 94.5 to 97.5 °C. (B) Concentration dependence of A65S thermal denaturation, in the presence of 2.4 M GuHCl. Protein concentrations were 0.9 (○) and 9 μM (●), and the T_m shifted from 92.5 to 98 °C. (C) Concentration dependence of R19A thermal denaturation, in the presence of 2.4 M GuHCl. Protein concentrations were 0.9 (○) and 9 μM (●).

is 113 ± 1 °C (inset of Figure 8). Figure 9 shows the dependence of the denaturation temperature of BldR2 and its mutants, A65S and R19A, in the presence of 2.4 M GuHCl at two different protein concentrations. For BldR2 and A65S, we observed a shift in the denaturation temperature in a single sigmoid-shaped curve indicative of a two-state dimeric unfolding process (Figure 9A,B), whereas thermal denaturation curves of R19A confirm the concentration

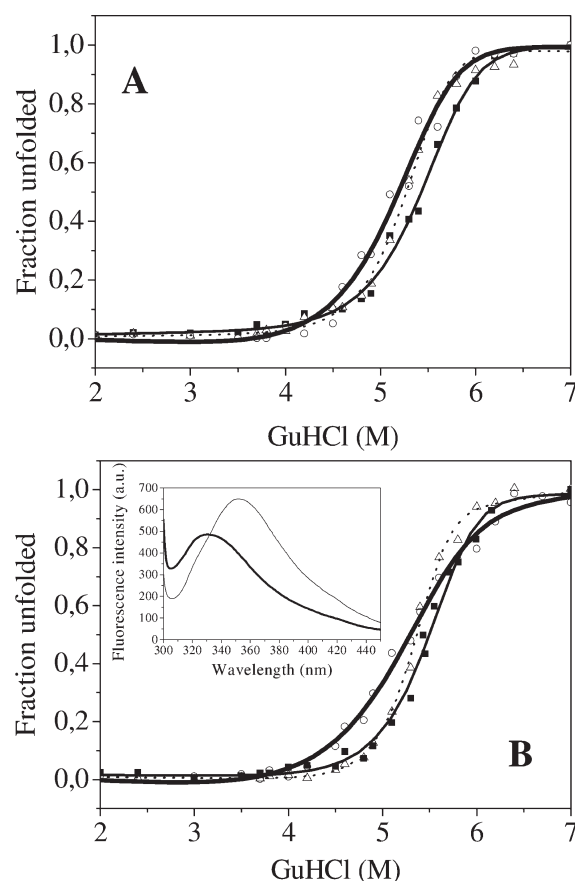


Figure 10. (A) GuHCl-induced unfolding curves of BldR2 (■), A65S (△), and R19A (○), obtained by recording the molar ellipticity at 230 nm and 25 °C. (B) Change in the fluorescence intensity at 336 nm and 25 °C. The lines are the best fits of the curves performed as described in Materials and Methods (inset of Figure 5B). Fluorescence emission spectra of native BldR2 in buffer solution (pH 7.5) (○) and in the presence of 7 M GuHCl (●) at 25 °C.

dependence but show a biphasic sigmoidal shape with two inflection points (Figure 9C). The thermal denaturation temperature of A65S proved to be very similar to that of BldR2 (111 ± 1 °C). Instead, the biphasic curve of R19A prevented us from correctly estimating the denaturation temperature and strongly suggested a more complex thermal unfolding process (Figure S4 of the Supporting Information). Because thermal denaturation of BldR2 and the mutants is not fully reversible under any of the investigated experimental conditions, no further thermodynamic analysis was performed.

Denaturant-Induced Unfolding. The conformational stability of BldR2 and its mutants against the denaturing action of GuHCl has been investigated at 25 °C in 20 mM sodium phosphate buffer (pH 7.5) by performing CD and fluorescence measurements. It is worth noting that urea cannot be used because BldR2 is extremely resistant to this denaturing agent. The transition curves obtained by recording the molar ellipticity at 230 nm (i.e., detecting the secondary structure stability) and that obtained by recording the fluorescence intensity (tertiary structure) are reported in panels A and B of Figure 10. The inset of Figure 10B shows the comparison of fluorescence emission spectra of BldR2 in the absence and presence of 7 M GuHCl when the protein is completely unfolded according to the

Table 2. Thermodynamic Parameters Obtained by the Analysis of the GuHCl-Induced Unfolding Curves of BldR2 Monitored by the Change in Molar Ellipticity at 230 nm ($[\theta]_{230}$) and of Fluorescence Intensity I_{336} at pH 7.5, 20 mM Sodium Phosphate Buffer, and 25 °C^a

sample	probe	$\Delta G_U(\text{H}_2\text{O})$ (kJ/mol)	m (kJ mol ⁻¹ M ⁻¹)	C_m (M)
BldR2	$[\theta]_{230}$	101 ± 5	13 ± 1	5.6
	I_{336}	106 ± 5	14 ± 1	5.6
A65S	$[\theta]_{230}$	116 ± 3	16 ± 1	5.4
	I_{336}	119 ± 4	16 ± 1	5.4
R19A	$[\theta]_{230}$	87 ± 6	11 ± 1	5.2
	I_{336}	82 ± 3	10 ± 1	5.2

^a Each figure is the average of the values calculated by the nonlinear regression procedure over three independent measurements. Errors are the standard deviations of the fits.

exposure of Trp and Tyr residues to the aqueous solvent. The different spectroscopic probes show superposable transition curves, indicating that both the secondary and tertiary structures are concurrently lost. The GuHCl-induced denaturation curves have a simple sigmoid shape for BldR2, A65S, and R19A, suggesting a cooperative two-state transition between folded dimers and unfolded monomers. The GuHCl-induced unfolding has proven to be reversible: fully unfolded samples showed recovery of spectroscopic features of the native protein after suitable dilution. The GuHCl concentrations at half-completion of the transition are 5.6, 5.4, and 5.2 M for BldR2, A65S, and R19A, respectively, highlighting the very high resistance to the denaturant. As the GuHCl-induced unfolding of BldR2 and A65S and R19A mutants is a reversible process, a thermodynamic analysis of the transitions can be performed. Linear extrapolation of ΔG_U (eq 3) yields the unfolding Gibbs energy in the absence of denaturant [$\Delta G_U(\text{H}_2\text{O})$] and the m value; all of the thermodynamic data of the unfolding transitions are listed in Table 2.

DISCUSSION

Because of their function in multidrug resistance and tolerance to highly toxic compounds, MarR transcriptional regulators have been intensely characterized in bacteria and archaea with the aim of understanding the molecular mechanisms of regulated response to such a stress. As many archaea live in hostile environments and are able to defend themselves from a wide variety of stress conditions, they represent an interesting model for determining their ability to survive under rapidly changing conditions. For such organisms, MarR-type regulators might be critical for their adaptation to particular habitats or lifestyles. Intriguingly, the exploration of homologous protein sequences from *Sulfolobales*, including different strains of *S. solfataricus*, *Sulfolobus islandicus*, *S. tokodaii*, and *Sulfolobus acidocaldarius*, revealed that, differently from the majority of bacteria, they all have at least two conserved MarR representatives, indicating the presence of several still uncharacterized regulatory systems involved in multiple-antibiotic resistance. MarR proteins also represent outstanding models for the study of protein stability for several reasons. (i) Many structures from distantly related organisms are available. (ii) They have a dimeric quaternary structure. (iii) They are able to bind DNA only as homodimers.

In this study, an archaeal MarR member, BldR2, was overexpressed in *E. coli* and purified to homogeneity. The protein forms a homodimer in solution and binds specifically to its own

promoter region with micromolar affinity, a value that is comparable to those other archaeal MarR members^{17,24,25} but is lower than those of many bacterial counterparts.

A primer extension analysis identified a transcription initiation site mapping 45 bp downstream of the first computationally predicted *Sso1082* codon³⁰ and according to transcriptome mapping.³¹ On the basis of our result, a promoter region was found containing consensus BRE-TATA sequences centered at positions −31 and −27 with respect to the transcription start site.

The binding site for BldR2 as determined by DNase I footprinting analysis was found in a region extending from BRE/TATA sequence to the transcription start site. Interestingly, the bound region contains an 8 bp pseudopalindromic sequence (BRE/TATA, AAACCTTTA) separated by 3 bp from an 8 bp perfect palindrome (TTTATAAA), suggesting that one BldR2 dimer could bind to each site. Hence, according to the location of the identified basal promoter elements and the BldR2 binding region, it can be proposed that the protein interferes with *Sso1082* transcription by competing with basal transcription factors.

Benzaldehyde and salicylate, known to act as effectors of different MarR proteins at millimolar concentrations, were able to weaken the interaction of BldR2 with its own promoter, suggesting that salicylate and to a lesser extent benzaldehyde could be ligands for BldR2. The low affinity of the effectors for the protein raises questions about their physiological relevance and indicates that other aromatic compounds could be the natural effectors. The in vitro binding results correlated with the in vivo induction of *BldR2* gene expression upon addition of aromatic drugs; the level of gene expression was also increased during the late-log growth phase. The derepression of *BldR2* both in the presence of aromatic compounds and in late-log growth phase supports a picture in which *BldR2* expression could be regulated by endogenous effectors derived from aromatic catabolic pathways. Hence, BldR2 in vivo would control regulatory mechanisms diverse from those regulated by BldR, which mainly works in the exponential growth phase, and in a different way.²⁵

Two single-point mutants, R19A and A65S, were also produced and characterized. The dimeric state of the mutants was confirmed by gel filtration experiments and thermal unfolding at different protein concentrations. The results of the thermodynamic characterization showed that BldR2 possesses a very stable dimeric conformation and that the mutants have also a very high resistance to both temperature and GuHCl denaturing action. In fact, the estimated denaturation temperature of BldR2 is in good agreement with that of the homologous ST1710 from *S. tokodaii*, obtained by differential scanning calorimetry¹⁷ and of the dimeric protein ORF56 from *S. islandicus*.⁴⁰ The monophasic temperature-induced unfolding curve of the wild-type protein suggested that the denaturation mechanism occurs in the absence of detectable equilibrium intermediates. This behavior is conserved in mutant A65S, suggesting that the mutation does not affect the protein global stability in solution. Interestingly, mutant R19A shows a biphasic sigmoid-shaped thermally induced unfolding curve. This finding suggests that, in this case, a more complex process occurs, going through the formation of stable intermediate species. On the basis of those results, it is tempting to speculate that residue R19 could be involved in important stabilizing interactions in the dimerization region of the protein, in a way that is reminiscent of what happens in the BldR protein. In fact, in BldR the conserved Arg19 residue of one monomer sets up a strong H bond (distance of 3 Å) with Tyr60 of the other

monomer.²⁴ The temperature-induced unfolding process is not reversible for all of the studied protein samples; this thermal irreversibility seems to be a common feature of MarR family members. In fact it was also found for the mesophilic HucR from *Deinococcus radiodurans*.⁴¹

The GuHCl-induced unfolding transitions of BldR2, monitored by CD in the far-UV region and fluorescence, are equivalent, indicating that both the protein secondary and tertiary structure are lost at the same time. This allows us to argue that the transition can be described by a two-state model for a dimer. The experimental values of C_m for BldR2 (5.6 M) and mutants A6SS (5.4 M) and R19A (5.2 M) indicate that the wild type is more resistant to the denaturant action than the mutants. The Gibbs energy of unfolding (101 kJ mol^{-1}) is in agreement with that of ORF56 from *S. islandicus* (85.1 kJ mol^{-1}).^{40,42} A closer analysis of ΔG_U and m values yields intriguing information about the structural properties of the wild-type protein and its mutants. In particular, even if A6SS has a C_m value lower than that of BldR2, its m value is higher and, as a consequence, ΔG_U values are very close. Because a higher m value is usually related to a larger change in the accessible surface area (ASA) upon unfolding, we can argue that the unfolded conformation of the A6SS mutant may be more exposed to the solvent.

Another interesting finding is represented by the denaturant-induced unfolding results of R19A that clearly show a simple sigmoid-shaped curve quite different from the biphasic profile obtained for the temperature unfolding curves. The result of the interpolation procedure, based on a simple two-state model, gives an overall lower stability of R19A compared to that of the wild-type protein; it is possible that this value represents an underestimation mainly caused by the choice of the model. In fact, it has been reported previously⁴³ that the presence of intermediate species in the chemically induced unfolding could not produce any dramatic perturbation in the monophasic unfolding curve, but in that case, the assumption of a two-state model would underestimate the Gibbs energy and the m value.⁴³ This conclusion fully agrees with the scenario that we previously anticipated on the basis of the thermal experiments in which Arg19 represents an important player in the stabilization of the dimer. Taken together, these results raise the possibility that a simple two-state model could not fully describe the BldR2 denaturation process. Possibly, the equilibrium unfolding of the BldR2 dimer could be described as an apparent two-state reversible reaction, in which unfolding and dissociation are coupled processes. This apparent two-state reaction seems to be confirmed by A6SS behavior, while for R19A, the unfolding and dissociation processes seem to be separated, because of the key position of the arginine residue in the dimerization domain.

The structure of the archaeal ST1710–DNA complex and a molecular model of the BldR–DNA complex showed that serine 65 was, in both cases, a critical residue for protein–DNA interaction acting as both a donor and an acceptor of H-bonds with its target DNA. Interestingly, the Ser68 in the bacterial OhrR of *Bacillus subtilis* was also found to make contacts with the major groove of the DNA molecule.⁴⁴ In our study, the characterization of the DNA binding properties of the A6SS mutant highlighted Ser65 as a key amino acid; in fact, this single-amino acid substitution was able not only to increase the extent of protein–DNA interaction by ~ 70 -fold but also to cause its binding to a further sequence that is a direct repeat located immediately upstream the TATA/BRE sequence in the *Sso1082*

promoter. In archaea, binding by transcriptional regulators to sequences in that position has been proven to be associated with a transcriptional activation.^{25,45}

This evidence suggests that the serine residue would allow the formation of an additional interaction with DNA responsible for an extension of the recognition sequence; this would cause a modification both in the binding mechanism and in the sequence specificity. On the basis of this observation, we can also hypothesize that this difference could correlate with the different proposed in vivo physiological roles of the two BldR proteins and depict A6SS as a protein intermediate between a repressor (BldR2) and an activator (BldR).

In conclusion, we propose for BldR2 a role in its autoregulation, but further analyses are required to understand the overall mechanism of regulation by BldR2, which would include a deeper investigation of the in vivo function of multiple transcription start sites, a genome-wide identification of target genes and natural ligands, and the analysis of *BldR*[−] and *BldR2*[−] mutant cells. Furthermore, this study highlights the idea that the identification of key residues involved in dimer stability may contribute to our understanding of the structural–functional relationship in the MarR family, because it is known that mutations in the dimerization domain are critical for the transcriptional regulation of MarR members.¹⁹ Moreover, given that mutations in the DNA binding domain can increase antibiotic tolerance,⁴⁶ we also suggest that knowledge of amino acids involved in DNA recognition may provide a remarkable starting point for the design of engineered MarR regulators acting as innovative therapeutic tools.

■ ASSOCIATED CONTENT

S Supporting Information. A densitometric analysis of a Northern blot, gel filtration profiles, far-UV CD spectra, and thermal denaturation curves at different GuHCl concentrations of A6SS and R19A. This material is available free of charge via the Internet at <http://pubs.acs.org>.

■ AUTHOR INFORMATION

Corresponding Author

*Phone: +39081679167 (G.F.) or +39081674255 (P.D.V.). Fax: +39081679053. E-mail: fiogabri@unina.it (G.F.) or pompea.delvecchio@unina.it (P.D.V.).

Funding Sources

This work was supported by MIUR-PRIN (Ministero dell'Istruzione, dell'Università e della Ricerca Scientifica-Progetti di Ricerca di Interesse Nazionale) (Grant 2008, CUP: E61J100000200001). The Centro Interdipartimentale di Metodologie Chimico Fisiche (CIMCF, University of Naples Federico II) is gratefully acknowledged for providing the spectrometers. L.M. is a fellow of the European Molecular Biology Organization (EMBO).

■ ACKNOWLEDGMENT

We acknowledge Dr. Raffaele Ronca for helpful discussion, Dr. Patrizia Contursi for critically reading the manuscript, and Dr. Omri Wurtzel for precious explanations about *S. solfataricus* transcriptomic data.

■ ABBREVIATIONS

BldR2, MarR member from *S. solfataricus*; IPTG, isopropyl 1-thio- β -D-galactopyranoside; SDS–PAGE, sodium dodecyl sulfate–polyacrylamide gel electrophoresis; EMSA, electrophoretic mobility shift assay; GuHCl, guanidine hydrochloride; CD, circular dichroism.

■ REFERENCES

- (1) Wheelis, M. L., Kandler, O., and Woese, C. R. (1992) On the nature of global classification. *Proc. Natl. Acad. Sci. U.S.A.* 89, 2930–2934.
- (2) Grabowski, B., and Kelman, Z. (2003) Archeal DNA replication: Eukaryal proteins in a bacterial context. *Annu. Rev. Microbiol.* 57, 487–516.
- (3) Verhees, C. H., Kengen, S. W., Tuininga, J. E., Schut, G. J., Adams, M. W., De Vos, W. M., and Van Der Oost, J. (2003) The unique features of glycolytic pathways in Archaea. *Biochem. J.* 375, 231–246.
- (4) van de Vossenberg, J. L., Driessen, A. J., and Konings, W. N. (1998) The essence of being extremophilic: The role of the unique archaeal membrane lipids. *Extremophiles* 2, 163–170.
- (5) Geiduschek, E. P., and Ouhammouch, M. (2005) Archaeal transcription and its regulators. *Mol. Microbiol.* 56, 1397–1407.
- (6) Hirata, A., Klein, B. J., and Murakami, K. S. (2008) The X-ray crystal structure of RNA polymerase from Archaea. *Nature* 451, 851–854.
- (7) Ramos, A., Raven, N., Sharp, R. J., Bartolucci, S., Rossi, M., Cannio, R., Lebbink, J., Van Der Oost, J., De Vos, W. M., and Santos, H. (1997) Stabilization of Enzymes against Thermal Stress and Freeze-Drying by Mannosylglycerate. *Appl. Environ. Microbiol.* 63, 4020–4025.
- (8) Jaenicke, R., and Zavodszky, P. (1990) Proteins under extreme physical conditions. *FEBS Lett.* 268, 344–349.
- (9) Razvi, A., and Scholtz, J. M. (2006) Lessons in stability from thermophilic proteins. *Protein Sci.* 15, 1569–1578.
- (10) Pedone, E., Bartolucci, S., and Fiorentino, G. (2004) Sensing and adapting to environmental stress: The archaeal tactic. *Front. Biosci.* 9, 2909–2926.
- (11) Schelet, J., Drozda, M., Dixit, V., Dillman, A., and Blum, P. (2006) Regulation of mercury resistance in the crenarchaeote *Sulfolobus solfataricus*. *J. Bacteriol.* 188, 7141–7150.
- (12) Limauro, D., Pedone, E., Pirone, L., and Bartolucci, S. (2006) Identification and characterization of 1-Cys peroxiredoxin from *Sulfolobus solfataricus* and its involvement in the response to oxidative stress. *FEBS J.* 273, 721–731.
- (13) Wilkinson, S. P., and Grove, A. (2006) Ligand-responsive transcriptional regulation by members of the MarR family of winged helix proteins. *Curr. Issues Mol. Biol.* 8, 51–62.
- (14) Newberry, K. J., Fuangthong, M., Panmanee, W., Mongkolsuk, S., and Brennan, R. G. (2007) Structural mechanism of organic hydroperoxide induction of the transcription regulator OhrR. *Mol. Cell* 28, 652–664.
- (15) Alekshun, M. N., Levy, S. B., Mealy, T. R., Seaton, B. A., and Head, J. F. (2001) The crystal structure of MarR, a regulator of multiple antibiotic resistance, at 2.3 Å resolution. *Nat. Struct. Biol.* 8, 710–714.
- (16) Perera, I. C., and Grove, A. (2010) Molecular mechanisms of ligand-mediated attenuation of DNA binding by MarR family transcriptional regulators. *J. Mol. Cell Biol.* 2, 243–254.
- (17) Kumarevel, T., Tanaka, T., Umehara, T., and Yokoyama, S. (2009) ST1710-DNA complex crystal structure reveals the DNA binding mechanism of the MarR family of regulators. *Nucleic Acids Res.* 37, 4723–4735.
- (18) Perera, I. C., Lee, Y. H., Wilkinson, S. P., and Grove, A. (2009) Mechanism for attenuation of DNA binding by MarR family transcriptional regulators by small molecule ligands. *J. Mol. Biol.* 390, 1019–1029.
- (19) Okada, N., Oi, Y., Takeda-Shitaka, M., Kanou, K., Umeyama, H., Haneda, T., Miki, T., Hosoya, S., and Danbara, H. (2007) Identification of amino acid residues of *Salmonella* SlyA that are critical for transcriptional regulation. *Microbiology (Reading, U.K.)* 153, 548–560.

- (20) Miyazono, K., Tsujimura, M., Kawarabayasi, Y., and Tanokura, M. (2007) Crystal structure of an archaeal homologue of multidrug resistance repressor protein, EmrR, from hyperthermophilic archaea *Sulfolobus tokodaii* strain 7. *Proteins* 67, 1138–1146.
- (21) Kumarevel, T., Tanaka, T., Nishio, M., Gopinath, S. C., Takio, K., Shinkai, A., Kumar, P. K., and Yokoyama, S. (2008) Crystal structure of the MarR family regulatory protein, ST1710, from *Sulfolobus tokodaii* strain 7. *J. Struct. Biol.* 161, 9–17.
- (22) Saridakis, V., Shahinas, D., Xu, X., and Christendat, D. (2008) Structural insight on the mechanism of regulation of the MarR family of proteins: High-resolution crystal structure of a transcriptional repressor from *Methanobacterium thermoautotrophicum*. *J. Mol. Biol.* 377, 655–667.
- (23) Okada, U., Sakai, N., Yao, M., Watanabe, N., and Tanaka, I. (2006) Structural analysis of the transcriptional regulator homolog protein from *Pyrococcus horikoshii* OT3. *Proteins* 63, 1084–1086.
- (24) Di Fiore, A., Fiorentino, G., Vitale, R. M., Ronca, R., Amodeo, P., Pedone, C., Bartolucci, S., and De Simone, G. (2009) Structural analysis of BldR from *Sulfolobus solfataricus* provides insights into the molecular basis of transcriptional activation in archaea by MarR family proteins. *J. Mol. Biol.* 388, 559–569.
- (25) Fiorentino, G., Ronca, R., Cannio, R., Rossi, M., and Bartolucci, S. (2007) MarR-like transcriptional regulator involved in detoxification of aromatic compounds in *Sulfolobus solfataricus*. *J. Bacteriol.* 189, 7351–7360.
- (26) Fiorentino, G., Ronca, R., and Bartolucci, S. (2009) A novel *E. coli* biosensor for detecting aromatic aldehydes based on a responsive inducible archaeal promoter fused to the green fluorescent protein. *Appl. Microbiol. Biotechnol.* 82, 67–77.
- (27) Brock, T. D., Brock, K. M., Belly, R. T., and Weiss, R. L. (1972) *Sulfolobus*: A new genus of sulfur-oxidizing bacteria living at low pH and high temperature. *Arch. Mikrobiol.* 84, 54–68.
- (28) Cannio, R., Fiorentino, G., Rossi, M., and Bartolucci, S. (1999) The alcohol dehydrogenase gene: Distribution among *Sulfolobales* and regulation in *Sulfolobus solfataricus*. *FEMS Microbiol. Lett.* 170, 31–39.
- (29) Limauro, D., Falcitatore, A., Basso, A. L., Forlani, G., and De Felice, M. (1996) Proline biosynthesis in *Streptococcus thermophilus*: Characterization of the proBA operon and its products. *Microbiology (Reading, U.K.)* 142 (Part 11), 3275–3282.
- (30) She, Q., Singh, R. K., Confalonieri, F., Zivanovic, Y., Allard, G., Awayez, M. J., Chan-Weiher, C. C., Clausen, I. G., Curtis, B. A., De Moors, A., Erauso, G., Fletcher, C., Gordon, P. M., Heikamp-de Jong, I., Jeffries, A. C., Kozera, C. J., Medina, N., Peng, X., Thi-Ngoc, H. P., Redder, P., Schenk, M. E., Theriault, C., Tolstrup, N., Charlebois, R. L., Doolittle, W. F., Duguet, M., Gaasterland, T., Garrett, R. A., Ragan, M. A., Sensen, C. W., and Van der Oost, J. (2001) The complete genome of the crenarchaeon *Sulfolobus solfataricus* P2. *Proc. Natl. Acad. Sci. U.S.A.* 98, 7835–7840.
- (31) Wurtzel, O., Sapra, R., Chen, F., Zhu, Y., Simmons, B. A., and Sorek, R. (2010) A single-base resolution map of an archaeal transcriptome. *Genome Res.* 20, 133–141.
- (32) Notredame, C., Higgins, D. G., and Heringa, J. (2000) T-Coffee: A novel method for fast and accurate multiple sequence alignment. *J. Mol. Biol.* 302, 205–217.
- (33) Lambert, C., Leonard, N., De Bolle, X., and Depiereux, E. (2002) ESyPred3D: Prediction of proteins 3D structures. *Bioinformatics* 18, 1250–1256.
- (34) Bradford, M. M. (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* 72, 248–254.
- (35) Gill, S. C., and von Hippel, P. H. (1989) Calculation of protein extinction coefficients from amino acid sequence data. *Anal. Biochem.* 182, 319–326.
- (36) Compton, L. A., and Johnson, W. C., Jr. (1986) Analysis of protein circular dichroism spectra for secondary structure using a simple matrix multiplication. *Anal. Biochem.* 155, 155–167.
- (37) Whitmore, L., and Wallace, B. A. (2004) DICHROWEB, an online server for protein secondary structure analyses from circular dichroism spectroscopic data. *Nucleic Acids Res.* 32, W668–W673.

- (38) Pace, C. N. (1986) Determination and analysis of urea and guanidine hydrochloride denaturation curves. *Methods Enzymol.* 131, 266–280.
- (39) Tanford, C. (1970) Protein denaturation. C. Theoretical models for the mechanism of denaturation. *Adv. Protein Chem.* 24, 1–95.
- (40) Zeeb, M., Lipps, G., Lilie, H., and Balbach, J. (2004) Folding and association of an extremely stable dimeric protein from *Sulfolobus islandicus*. *J. Mol. Biol.* 336, 227–240.
- (41) Wilkinson, S. P., and Grove, A. (2004) HucR, a novel uric acid-responsive member of the MarR family of transcriptional regulators from *Deinococcus radiodurans*. *J. Biol. Chem.* 279, 51442–51450.
- (42) Luke, K. A., Higgins, C. L., and Wittung-Stafshede, P. (2007) Thermodynamic stability and folding of proteins from hyperthermophilic organisms. *FEBS J.* 274, 4023–4033.
- (43) Soulages, J. L. (1998) Chemical denaturation: Potential impact of undetected intermediates in the free energy of unfolding and m-values obtained from a two-state assumption. *Biophys. J.* 75, 484–492.
- (44) Hong, M., Fuangthong, M., Helmann, J. D., and Brennan, R. G. (2005) Structure of an OhrR-ohrA operator complex reveals the DNA binding mechanism of the MarR family. *Mol. Cell* 20, 131–141.
- (45) Kessler, A., Sezonov, G., Guijarro, J. I., Desnoues, N., Rose, T., Delepierre, M., Bell, S. D., and Prangishvili, D. (2006) A novel archaeal regulatory protein, Stal, activates transcription from viral promoters. *Nucleic Acids Res.* 34, 4837–4845.
- (46) Kim, J. Y., Inaoka, T., Hirooka, K., Matsuoka, H., Murata, M., Ohki, R., Adachi, Y., Fujita, Y., and Ochi, K. (2009) Identification and characterization of a novel multidrug resistance operon, mdtRP (yusOP), of *Bacillus subtilis*. *J. Bacteriol.* 191, 3273–3281.